KEVIN KELLY, OLIVER SCHULTE, VINCENT HENDRICKS

RELIABLE BELIEF REVISION

ABSTRACT. Philosophical logicians proposing theories of rational belief revision have had little to say about whether their proposals assist or impede the agent's ability to reliably arrive at the truth as his beliefs change through time. On the other hand, reliability is the central concern of formal learning theory. In this paper we investigate the belief revision theory of Alchourrón, Gärdenfors and Makinson from a learning theoretic point of view.

1. CONSERVATISM AND RELIABILITY

There are two fundamentally different perspectives on the study of belief revision. A *conservative* methodologist seeks to minimize the damage done to his current beliefs by new information. A *reliabilist*, on the other hand, seeks to find the truth whatever the truth might be. Both aims support hypothetical imperatives about how inquiry ought to proceed, imperatives that might be considered principles of inductive rationality. It would seem that there is some tension between the two perspectives. The conservative sentiment is to smooth over the effects of new information, whereas reliable inquiry may require more radical changes. The inductive leap from a hundred black ravens to the universal generalization that all ravens are black is not conservative. Nor was Copernicus' revolutionary rejection of conservative tinkering within the Ptolemaic system. Conservatism and reliabilism are reflected in two equally distinctive logical perspectives on the problem of belief revision. Conservatism is the motivation behind the theory of belief revision proposed by Alchourrón, Makinson, and Gärdenfors (AGM). Reliabilism is the principal concern of formal learning theory.[1]

Although the AGM approach is not motivated by reliability considerations, its authors seem to hope for the best.

> Even if the analysis of belief systems presented in this book does not depend on any account of the relation between such systems and reality, I still believe that rational belief systems to a large extent mirror an actual world. (Gärdenfors, 1988, p.19)

But if this happy situation obtains, it must have been brought about by the process of belief revision. Therefore, an important question for any proposed theory of rational belief revision is whether its norms assist or interfere with an agent's ability to arrive at the truth as evidence accumulates. Indeed, when putative principles of inductive rationality can be shown to prevent one from being as reliable as one otherwise might have been, we are strongly inclined to side with reliability rather than with the principles.[2]

---

[1] For general introductions, cf. (Osherson *et al.*, 1986) and (Kelly, 1995).

[2] Hilary Putnam (Putnam, 1963) responded to Rudolf Carnap's theory of confirmation in just this way.

In this paper, we provide a learning theoretic analysis of the effects of the AGM axioms on reliabile inductive inquiry. We consider a variety of ways in which the AGM theory can be interpreted as constraining the course of inductive inquiry. A representation of the reliable AGM methods is established.

2. THEORY DISCOVERY

We will adopt a fairly simple and abstract construal of reliable theory discovery (Kelly, 1995). We assume that the scientist is studying some system with discrete observable states that may be encoded by natural numbers, so that in the limit the scientist receives an infinite stream ε of code numbers. Let $\mathcal{N}$ denote the set of all such data streams.

An *empirical proposition* is is a proposition whose truth or falsity depends only on the data stream. We therefore identify theories and hypotheses with sets of data streams. Let Λ be a countable set of empirical propositions. We may think of Λ as the set of all propositions we want to find the truth about. We assume only that Λ is closed under complementation. A Λ-theory is defined to be any result of intersecting some collection of propositions drawn from Λ. The complete Λ-theory of a data stream ε is just the intersection of all propositions in Λ that contain ε. Set-theoretic relations among propositions represent logical connectives and relations in the usual way: entailment is set inclusion, inconsistency is disjointness, conjunction is finite intersection, and so forth.

Let e be a finite data sequence. The empirical proposition expressed by e is just the set of all infinite data streams extending e, which will be denoted by $[e]$. We refer to such propositions as *fans*. $e^\frown q$ denotes the result of concatenating datum q onto the end of finite sequence e.

A *theory discovery method* δ takes finite data sequences as inputs and outputs propositions. In particular, δ produces an initial conjecture $\delta(\emptyset)$ on the empty sequence $\emptyset$. We don't require that the output proposition be in Λ. We will think of the method as reading ever longer initial segments of some infinite data stream ε. It is therefore convenient to let $\varepsilon|n$ denote the initial segment of ε of length n: i.e., $(\varepsilon(0), ..., \varepsilon(n-1))$. Now we define two concepts of reliable theory discovery.[3] Uniform discovery requires that the method be guaranteed to eventually produce only consistent conjectures entailing the complete Λ truth. Non-uniform or piecemeal discovery requires that each proposition in Λ has its truth value eventually settled by the method's conjectures, but there may be no time by which all propositions in Λ have their truth values settled.

Definition 1 *Reliable Theory Discovery*

[3] (Kelly and Glymour, 1990) and (Kelly, 1995) define these concepts in a first-order logical setting.

δ uniformly discovers *the complete* Λ *truth* $\Longleftrightarrow$ *for each data stream* ε *there is a time such that for each later time* m, $\delta(\varepsilon|m)$ *is consistent and entails the complete* Λ*-theory of* ε.

δ piecemeal discovers *the complete* Λ *truth* $\Longleftrightarrow$ *for each data stream* ε *and for each proposition* $\mathcal{P}$ *in* Λ *there is a time such that for each later time* m, $\delta(\varepsilon|m)$ *entails* $\mathcal{P}$ *just in case* $\mathcal{P}$ *is true of* ε.

Clearly, uniform success entails piecemeal success. Also, there is no guarantee that the successive theories produced by a piece-meal discovery method get ever more "verisimilar". A piece-meal method is permitted to add as much new falsehood as it pleases at each stage, so long as more and more propositions have their truth values correctly settled.

A few examples may clarify the difference between the two reliability concepts. Let Λ_0 be the set of all fans and their complements. It is trivial to find the complete Λ_0 truth in a piecemeal fashion simply by returning $[e]$ on input data sequence e. But no method can do so uniformly, since for each data stream ε, $\{\varepsilon\}$ is the complete Λ_0-truth, and whereas there are uncountably many such singleton theories, the range of each discovery method is countable, so most such theories cannot even be conjectured.

Even when there are only countably many distinct Λ-complete theories, piecemeal success may be possible when uniform success is not. Let the proposition $\mathcal{H}_n$ say that 0 occurs at stage n and forever after. Let Λ_1 be the set of all such propositions and their negations. Λ_1 requires piecemeal solutions to perform nontrivial inductive inferences, since each hypothesis $\mathcal{H}_n$ is a claim about the unbounded future. This time, there are only countably many distinct Λ_1-complete theories[4]. To succeed piecemeal, let method β conjecture $\mathcal{H}_{n+1} \cap \overline{\mathcal{H}_n}$ when the last non-0 occurring in the data occurs at position n. Nonetheless, uniform success is still impossible. For suppose for reductio that some method δ succeeds uniformly. Then a wily demon can present 0,0,0,... until δ's conjecture entails $\mathcal{H}_0$, which it must eventually do on that data stream. Then a 1 is fed (e.g., at stage m) followed by all 0s until δ's conjecture entails $\mathcal{H}_{m+1}$, which it must eventually do on that data stream, etc. The data stream ε so presented never stabilizes to 0, so δ produces infinitely many conjectures inconsistent with the complete Λ_1-theory of ε.

The same argument shows that no piecemeal solution could possibly succeed by eventually producing only true hypotheses, since the construction forces an arbitrary piecemeal solution to produce infinitely many false conjectures. Hence, piecemeal success is sometimes possible only if infinitely many false conjectures are produced. On the other hand, whenever uniform success is possible, it can be achieved by a method whose conjectures are eventually true. Given successful method δ, produce whatever δ says if it is not a consis-

[4] Such a theory either says exactly when the data stream stabilizes to 0 or says that the data stream never stabilizes to 0.

tent proposition entailing some Λ-complete theory and produce the (unique) Λ-complete theory entailed by the current conjecture of δ otherwise.

3. BELIEF REVISION

A belief revision operator $+$ takes a theory $\mathcal{T}$ and a proposition $\mathcal{P}$ and returns a revised theory $\mathcal{T}'$. Then we write

$$\mathcal{T} + \mathcal{P} = \mathcal{T}'$$

In our setting of empirical propositions, the AGM axioms amount to the following:

(AGM 1) $\mathcal{T} + \mathcal{P} \subseteq \mathcal{P}$.
(AGM 2) If $\mathcal{P} \neq \emptyset$, then $\mathcal{T} + \mathcal{P} \neq \emptyset$.
(AGM 3) If $\mathcal{P} \cap \mathcal{T} \neq \emptyset$, then $\mathcal{T} + \mathcal{P} = \mathcal{P} \cap \mathcal{T}$.
(AGM 4) If $\mathcal{P} \cap (\mathcal{T} + \mathcal{Q}) \neq \emptyset$, then $\mathcal{T} + (\mathcal{P} \cap \mathcal{Q}) = \mathcal{P} \cap (\mathcal{T} + \mathcal{Q})$.

We employ an elegant representation of these requirements due to A. Grove (1988). A *Grove system* for proposition $\mathcal{K}$ is a collection Γ of subsets of $\mathcal{N}$ such that

1. Γ is nested (i.e., totally ordered by $\subseteq$),
2. For each consistent proposition $\mathcal{P}$ there is a unique, $\subseteq$-minimal element of Γ whose intersection with $\mathcal{P}$ is nonempty.
3. $\mathcal{N} \in \Gamma$.
4. $\mathcal{K}$ is the least member of Γ (with respect to $\subseteq$).

Let $\Gamma_{\mathcal{P}}$ denote the least element of Γ whose intersection with $\mathcal{P}$ is nonempty. This is well-defined by the second condition on Grove systems. Let $\Gamma(\mathcal{P}) = \Gamma_{\mathcal{P}} \cap \mathcal{P}$. Then we have:

Proposition 1 (Grove 1988) *For each* AGM *belief revision operator* $+$ *and for each proposition* $\mathcal{K}$ *there is a Grove system* $\Gamma^{\mathcal{K}}$ *such that for each proposition* $\mathcal{P}$

$$\Gamma^{\mathcal{K}}(\mathcal{P}) = \mathcal{K} + \mathcal{P}.$$

For each proposition $\mathcal{K}$*, let* $\Gamma^{\mathcal{K}}$ *be a Grove system for* $\mathcal{K}$*. Then there exists a unique* AGM *belief revision operator* $+$ *such that for each* $\mathcal{K}, \mathcal{P}$

$$\Gamma^{\mathcal{K}}(\mathcal{P}) = \mathcal{K} + \mathcal{P}.$$

4. INDUCTION AS BELIEF REVISION

There are several ways in which the AGM theory of belief revision might interact with inductive methodology. An ambitious approach views induction as nothing but the revision of one's previous beliefs in accordance with the total evidence available. On this proposal, one's revision operator uniquely determines one's inductive method once an initial theory is specified. There are two ways in which this might happen. According to the first, inquiry begins with a fixed theory $\mathcal{K}$ and this theory is repeatedly updated on increasing data. We will refer to this as *repeated* revision. Repeated revision of $\mathcal{K}$ by a belief revision operator $+$ uniquely determines a discovery method δ according to the following relation:

1. $\delta(\emptyset) = \mathcal{K}$.
2. $\delta(e^\frown q) = \delta(\emptyset) + [e^\frown q]$.

Then we say that δ is *represented repetitively* by $\mathcal{K}$ and $+$. If δ is repetitively represented by some AGM operator starting with $\mathcal{K}$, then we say that δ is *repetitively* AGM. A second proposal is that it is always one's current theory that is revised as the data comes in. We will refer to this as *sequential* revision. Sequential revision starting with $\mathcal{K}$ uniquely determines a discovery method as follows:

1. $\delta(\emptyset) = \mathcal{K}$.
2. $\delta(e^\frown q) = \delta(e) + [e^\frown q]$.

Then we say that δ is *represented sequentially* by $\mathcal{K}$ and $+$.[5] If δ is sequentially represented by some AGM operator starting with $\mathcal{K}$, then we say that δ is *sequentially* AGM. The two concepts are equivalent for a given operator $+$ if the operator satisifes the following principle:

$$\mathcal{K} + [e^\frown q] = (\mathcal{K} + [e]) + [e^\frown q].$$

We refer to this as the *irrelevance of earlier subdata* (IES) principle. The IES principle does not follow from the AGM axioms, however, so repeated revision by a given operator is not the same as sequential revision by that operator.

It follows immediately from the AGM axioms that a sequentially AGM discovery method has various properties. First, it is *data retentive*, in the sense that its current conjecture always entails the data. Second, it is *consistent*, in the sense that its current conjecture is always consistent with its current data. Third, it is *stubborn* in the sense that its current conjecture always entails its preceding conjecture unless the preceding conjecture is refuted. Finally, it is *timid* in the sense that it only adds the data to its previous conjecture unless its preceding conjecture is refuted.

In fact, these properties characterize both the sequentially and the repetitively AGM discovery methods, so although it matters for a particular revision operator whether it is interpreted sequentially or repetitively, the same methods are representable either way. We may now refer to discovery methods that are data retentive, consistent, stubborn and timid simply as AGM methods.

[5] Both proposals are developed in (Martin and Osherson, 1995). Sequential revision is there referred to as 'iterated' revision.

Proposition 2 *The following statements are equivalent:*

1. *δ is data retentive, consistent, stubborn, and timid.*
2. *δ is sequentially* AGM.
3. *δ is repetitively* AGM.

The proof is given in the appendix.

Data retentiveness is a fairly tame requirement for an ideal[6] logic of discovery: one may as well remember the data. Consistency is as well: while we produce contradictions, we couldn't possibly be on the path to the truth.[7] Stubbornness and timidity are another matter. Consider a stubborn scientist whose only belief is that a given coin will come up heads at least once. Suppose it never comes up heads. He can never retract his belief unless he adds other "auxilliary hypotheses" so that his whole belief set is eventually refuted, affording him an opportunity to retract the original false belief. But if this scientist is also timid, then he never gets an opportunity to add any such auxilliary hypotheses, so he is hopelessly frozen for eternity in a false belief. Since there is no problem finding the truth about this hypothesis (conjecture its negation until one head is observed and then conjecture it forever after), it is clear that there are some initial belief sets from which no AGM operator can recover, either in the repetitive or in the sequential sense.

So for some initial belief states, sequential and repetitive AGM updating does stand in the way of finding the truth. The question remaining is whether some AGM method with the right kind of initial belief set can succeed whenever success is possible. The preceding discussion already provides a clue: if one always produces refutable conjectures, one never ends up in the awkward situation of not getting an opportunity to add the right kind of auxilliary hypotheses. A trivial way to ensure this is to always produce an empirically complete hypothesis (i.e., a hypothesis that singles out a unique data stream).

Proposition 3 *Suppose that the complete Λ truth is uniformly [piecemeal] discoverable. Then the complete Λ truth is uniformly [piecemeal] discoverable by an* AGM *method.*

Proof: Let δ uniformly discover the complete Λ truth. Let $\phi(\mathcal{K}, e)$ denote the choice of some data stream in $\mathcal{K} \cap [e]$ if $\mathcal{K} \cap [e] \neq \emptyset$ and let $\phi(\mathcal{K}, e)$ denote an arbitrary data stream in $[e]$ otherwise. Then define

$$\gamma(\emptyset) = \{\phi(\delta(\emptyset), \emptyset)\};$$

$$\gamma(e^\frown q) = \gamma(e) \text{ if } \gamma(e) \cap [e^\frown q] \neq \emptyset;$$

$$\gamma(e^\frown q) = \{\phi(\delta(e^\frown q), e^\frown q)\} \text{ otherwise.}$$

[6] By "ideal" we mean "abstracted from all computability considerations".

[7] On the other hand, computable inquiry can be very severely restricted by imposing the consistency requirement (Kelly and Schulte, 1995a; Kelly and Schulte, 1995b).

By the definition of ϕ, γ is consistent and data-retentive. By the second clause of γ's definition, γ is stubborn and timid. So by proposition 2, γ is an AGM method. Since δ uniformly discovers the complete Λ truth, we have that on each data stream ε, there is a stage n after which δ always produces consistent conjectures entailing the complete Λ truth. Suppose that γ stabilizes to some conjecture $\{\tau\}$ along ε. Then $\tau = \varepsilon$, so γ succeeds uniformly. Suppose then that γ never stabilizes to any conjecture along ε. Then γ selects a new conjecture after stage n. Thereafter, γ's conjecture always entails the complete Λ truth because δ's conjecture does.

Now suppose δ discovers the complete Λ truth in the piecemeal sense. Construct γ just as before. Again, if γ happens to stabilize to some conjecture on ε, then γ succeeds uniformly. So suppose γ never stabilizes to a conjecture. For each $\mathcal{H} \in \Lambda$, there is a stage n by which δ's conjecture entails $\mathcal{H}$ if $\mathcal{H}$ is true of ε and entails $\overline{\mathcal{H}}$ otherwise. So there is a stage m after n such that each successive conjecture of γ also has this property. □

An AGM operator $+$ is said to be *maxichoice* just in case $\mathcal{K} + \mathcal{A}$ is the result of adding $\mathcal{A}$ to some minimal superset $\mathcal{S}$ of $\mathcal{K}$ consistent with $\mathcal{A}$, in the sense that any proper subset of $\mathcal{S}$ that includes $\mathcal{K}$ would not be consistent with $\mathcal{A}$. The method γ constructed in the preceding proof can be represented by a maxichoice operator, since $\gamma(e^\frown q) \cup \gamma(e)$ is always a minimal superset of $\gamma(e)$ consistent with $[e^\frown q]$. This is the strongest sense of "minimal change" considered by Gärdenfors (and rejected for being too strong), but even it does not interfere with the possibility of reliable inquiry (although we have seen that the AGM axioms themselves interfere with reliability for agents starting out with the wrong kinds of beliefs).[8]

On the other hand, conservative intuitions about minimal belief change are hardly comforted by the piecemeal discovery method γ just constructed. γ leaps from one complete theory to another in order to ensure that it can change its mind when necessary without violating stubbornness and timidity. This raises the question whether we can say something in general about how strong the conjectures of a reliable AGM method must be. A few topological concepts are helpful. A *limit point* of an empirical proposition $\mathcal{K}$ is a data stream along which $\mathcal{K}$ is never refuted. The *empirical closure* of $\mathcal{K}$ (denoted $\mathrm{cl}(\mathcal{K})$) is the set of all limit points of $\mathcal{K}$ (i.e. the set of all data streams along which $\mathcal{K}$ is never refuted). $\mathcal{H}$ is *empirically decidable with certainty* or *empirically clopen* given $\mathcal{K}$ just in case the truth value of $\mathcal{H}$ is eventually decided by the data, assuming that $\mathcal{K}$ is true. That is, on each data stream in $\mathcal{K}$, there is a time after which the data together with $\mathcal{K}$ either entail $\mathcal{H}$ or entail the negation of $\mathcal{H}$. $\mathcal{H}$ is *empirically decidable with certainty by stage* n or *empirically* n*-clopen* given $\mathcal{K}$ just in case the truth value of $\mathcal{H}$ is determined by stage n, given that $\mathcal{K}$ is true. Now we have:

[8](Martin and Osherson, 1995, Prop. 23) show that if the notion of logical consequence is weakened, an initial belief set can be found that succeeds with any AGM operator in the repetitive sense. We do not consider weakened consequence relations here.

Proposition 4 *Let δ be a discovery method.*

1. *If δ is data-retentive, stubborn and timid and δ piecemeal discovers the complete Λ truth, then for each e, each proposition in Λ is empirically clopen given $cl(\delta(e))$.*
2. *If δ is data retentive, stubborn and timid and δ uniformly discovers the complete Λ truth, then for each e, there is an n such that each proposition in Λ is empirically n-clopen given $cl(\delta(e))$.*

Proof: (1) Suppose otherwise. Then for some e, and for some $\mathcal{H} \in \Lambda, \mathcal{H}$ is not clopen in $\mathrm{cl}(\delta(e))$. Hence, $(*)$ there is a data stream ε along which $\delta(e)$ is never refuted such that at no stage does the data read along ε entail either $\mathcal{H}$ or its complement. Since δ is data retentive, $\delta(e) \subseteq [e]$, so for some $n, e = \varepsilon|n$. Hence $\delta(\varepsilon|n) = \delta(e)$. But since δ is stubborn, timid and data retentive, for each $m \geq n, \delta(\varepsilon|m) = \delta(e) \cap [\varepsilon|m]$. So by $(*)$, δ does not piecemeal discover the complete Λ truth.

The proof of (2) is similar. □

The first condition is equivalent to saying that the topological boundary of each proposition in Λ must be excluded from each conjecture produced by δ. In a setting with first-order sentences instead of propositions, this would mean that each of δ's conjectures entails that each sentence in Λ is equivalent to a quantifier-free sentence (Kelly, 1995, Thm. 12.7). This is a striking requirement, but in the propositional setting it can always be accomplished with conjectures of arbitrarily high probability if probability is countably additive.[9]

5. RELIABILITY AS A REVISION THEORETIC AXIOM

AGM belief revision theory makes no reference to reliability. What would happen if we were to add as an AGM axiom that each operator be a reliable solution to a specified, solvable inductive inference problem Λ? It would be nice to have an exact representation, analogous to Grove's, of the class of all AGM operators that are reliable solutions to Λ. We have not yet found such a representation for the notions of reliability defined above. But it is much easier to obtain such a result if we modify our notions of reliability so as to require in addition, that on each data stream there is a time after which all the theories produced by δ are *true* (recall that uniform and piecemeal discovery are both possible even when no conjecture is true). Then we speak, respectively, of *truly* discovering the complete Λ truth, uniformly or in a piecemeal fashion.

Recall from section 2 that the complete Λ truth is truly uniformly discoverable just in case it is uniformly discoverable. On the other hand, true piecemeal discovery is properly easier than uniform discovery and piecemeal discovery is properly easier than true, piecemeal discovery.[10]

[9] Cf. (Kelly, 1995), proposition 13.15.

[10] The first noninclusion is witnessed by the example Λ_0 provided at the beginning of the paper. We have already seen that this inductive problem is not uniformly solvable. But the

Two more concepts are necessary in order to represent the "truly" reliable AGM methods. $\mathcal{K}$ is *empirically closed* just in case $\mathcal{K}$ contains all its limit points (i.e. just in case $\mathcal{K}$ is identical to its own empirical closure). Then $\mathcal{K}$ is guaranteed to be refuted eventually by the data if it is false, for otherwise there would be a data stream making $\mathcal{K}$ false along which $\mathcal{K}$ is never refuted (i.e., $\mathcal{K}$ is missing one of its limit points). Let Γ be a Grove system and let $\mathcal{S} \in \Gamma$. Let core($\mathcal{S}$) denote the union of all $\mathcal{R} \in \Gamma$ such that $\mathcal{R}$ is a proper subset of $\mathcal{S}$. Then we have:

Proposition 5 *Truly Reliable* AGM *Learners*

1. *δ is a repetitively* AGM *method that truly, piecemeal discovers the complete Λ truth $\iff$ there is a Grove system Γ such that*
   (a) $\forall e, \Gamma([e]) = \delta(e)$
   (b) $\forall \mathcal{S} \in \Gamma, core(\mathcal{S})$ *is empirically closed, and*
   (c) $\forall e, \forall \mathcal{H} \in \Lambda, \mathcal{H}$ *is empirically clopen in* $\Gamma([e])$.
2. *δ is a repetitively* AGM *method that truly, uniformly discovers the complete Λ truth $\iff$ there is a Grove system Γ such that*
   (a) $\forall e, \Gamma([e]) = \delta(e)$
   (b) $\forall \mathcal{S} \in \Gamma, core(\mathcal{S})$ *is empirically closed, and*
   (c) $\forall e, \exists n$ *such that* $\forall \mathcal{H} \in \Lambda$, $\mathcal{H}$ *is empirically n-clopen in* $\Gamma([e])$.

Proof of (1): ($\Longrightarrow$) suppose δ is a repetitively AGM method. Then by proposition 1, there is a Grove system Γ such that $\forall e, \Gamma([e]) = \delta(e)$, so we have 1a. 1c follows from 1a and proposition 4. Finally, suppose for reductio that 1b is false. Then let $\mathcal{S} \in \Gamma$ be such that some $\varepsilon \notin$ core($\mathcal{S}$) is a limit point of core($\mathcal{S}$). But then for each n, $\Gamma_{[\varepsilon|n]} \subseteq$ core($\mathcal{S}$), so $\varepsilon \notin \Gamma([\varepsilon|n]) = \delta([\varepsilon|n])$. Hence, δ produces infinitely many false conjectures along ε and so δ fails to truly piecemeal identify the complete Λ truth.

($\Longleftarrow$) Suppose there is a Grove system Γ satisfying 1a–1c. Let ε be given. Since Γ is a Grove system, there is a least element $\mathcal{S} \in \Gamma$ such that $\varepsilon \in \mathcal{S}$. So $\varepsilon \notin$ core($\mathcal{S}$). There is an n such that $[\varepsilon|n] \cap$ core($\mathcal{S}$) $= \emptyset$, else ε is a missing limit point of core($\mathcal{S}$), contrary to 1b. But since $\varepsilon \in \mathcal{S}$, for each $m \geq n, \Gamma([\varepsilon|m]) = \Gamma([\varepsilon|n]) \cap [\varepsilon|m]$. By 1c, we have that for each $\mathcal{H} \in \Lambda, \exists k \geq n. \forall j \geq k, \Gamma([\varepsilon|n]) \cap [\varepsilon|k]$ either entails $\mathcal{H}$ if $\varepsilon \in \mathcal{H}$ or entails $\overline{\mathcal{H}}$ otherwise. By 1a, δ truly piecemeal discovers the complete Λ truth.

The proof of (2) is similar. □

The requirement that the method stabilize to the truth undercuts the piecemeal strategy of succeeding by always producing a complete, false theory. We

trivial method that always repeats the current data succeeds truly in the nonuniform sense. The second noninclusion is witnessed by the example Λ_1 also presented at the beginning of the paper. It was shown above that this problem has a piecemeal solution. But the demonic argument against uniform solutions forces an arbitrary method to produce infinitely many false conjectures, so no true piecemeal solution is possible.

leave open the interesting question whether AGM methods can solve every problem solvable in the truly piecemeal sense.

6. ARBITRARY REVISIONS

So far, we have examined an ambitious interpretation of belief revision theory, in which one's belief revision operator uniquely determines one's inductive inferences. It has been shown that even on this interpretation, the AGM axioms do not restrict the reliability of inquiry, at last for ideal methods that needn't worry about computational limitations. On the other hand, AGM methods have some awkward properties such as timidity and stubbornness, and these requirements must be "steered around" to arrive at the truth.

But there is a more tempered interpretation of belief revision theory, according to which a revision operator is just a way of consistently forcing into one's beliefs whatever one decides to force into them for whatever reason. Say that a method is repetitively AGM *with arbitrary revisions* just in case it always updates its initial beliefs on the conjunction of the total data and some arbitrarily selected proposition. The definition of sequentially AGM methods with arbitrary revisions is similar, except that it is the initial belief set that is revised.

Definition 2 AGM *Learners With Arbitrary Revisions*

1. *δ is repetitively* AGM *with arbitrary revisions* $\iff$ *there is an* AGM *revision operator $+$ such that for all $e^\frown q$ there is an $\mathcal{A}$ such that* $\delta(e^\frown q) = \delta(\emptyset) + ([e^\frown q] \cap \mathcal{A})$.
2. *δ is sequentially* AGM *with arbitrary revisions* $\iff$ *there is an* AGM *revision operator $+$ such that for all $e^\frown q$ there is an $\mathcal{A}$ such that* $\delta(e^\frown q) = \delta(e) + ([e^\frown q] \cap \mathcal{A})$.

Neither concept implies stubbornness or timidity, since $\mathcal{A}$ can be chosen to be inconsistent with $\delta(\emptyset)$ (or respectively, with $\delta(e)$). Data retentiveness still follows. So do weakened versions of stubbornness and consistency. Stubbornness obliges δ to move forward until contradicted by the data. An 'internal' version of stubbornness requires δ to move forward until contradicted either by the data or by δ's new beliefs. Accordingly, δ is *internally stubborn in the repetitive [sequential] sense* just in case δ's current beliefs either entail its initial [current] beliefs or are inconsistent with them. Finally, say that δ is *quasi-consistent* just in case δ's beliefs are never both inconsistent with the data and consistent. Then we have:

Proposition 6 *Let δ be a discovery method. Then*

1. *δ is repetitively [sequentially]* AGM *with arbitrary revisions* $\iff$ *δ is data retentive and internally stubborn in the repetitive [sequential] sense.*

2. This remains true if we add the condition that δ is quasi-consistent.

Proof of 1, sequential case: ($\Longrightarrow$) Let δ be sequentially AGM with arbitrary revisions. Let $+$ be the AGM operator that witnesses this fact. Since δ is defined in terms of an AGM operator revising on at least the data, δ is data retentive. Now suppose that $\delta(e^\frown q)$ is consistent with $\delta(e)$. Since δ is sequentially AGM with arbitrary revisions, there is an $\mathcal{A}$ such that $\delta(e^\frown q) = \delta(e)+([e^\frown q]\cap\mathcal{A})$. By axiom (AGM 1), $\delta(e)+([e^\frown q]\cap\mathcal{A}) \subseteq ([e^\frown q]\cap\mathcal{A})$. Hence, $([e^\frown q]\cap\mathcal{A})$ is consistent with $\delta(e)$. So by axiom (AGM 3), $\delta(e)+([e^\frown q]\cap\mathcal{A}) = \delta(e) \cap ([e^\frown q] \cap \mathcal{A}) \subseteq \delta(e)$. Hence, $\delta(e^\frown q) \subseteq \delta(e)$.

($\Longleftarrow$) Let δ be data retentive and internally stubborn. For each proposition $\mathcal{A}$ let Grove system $\Gamma^{\mathcal{A}} = \{\mathcal{N}, \mathcal{A}\}$. Let $+$ be the operator represented by the collection of all such $\Gamma^{\mathcal{A}}$ according to proposition 1. Then we have

$$(*) \text{ if } \mathcal{S} \cap \mathcal{R} = \emptyset \text{ then } \mathcal{S} + \mathcal{R} = \mathcal{R}.$$

By data retentiveness, $\delta(e^\frown q) \subseteq [e^\frown q]$, so

$$(**)\ \delta(e^\frown q) \cap [e^\frown q] = \delta(e^\frown q).$$

By internal stubbornness in the sequential sense, there are two cases. Case 1: $\delta(e) \cap \delta(e^\frown q) = \emptyset$. Then by $(**)$ and $(*)$, $\delta(e) + ([e^\frown q] \cap \delta(e^\frown q)) = \delta(e) + \delta(e^\frown q) = \delta(e^\frown q)$.

Case 2: $\delta(e^\frown q) \subseteq \delta(e)$. Then by $(**)$, (AGM 3), and the case hypothesis, we have that $\delta(e) + ([e^\frown q] \cap \delta(e^\frown q)) = \delta(e) + \delta(e^\frown q) = \delta(e) \cap \delta(e^\frown q) = \delta(e^\frown q)$.

Proof of 1, repetitive case: The ($\Longrightarrow$) side is just as before, with $\delta(\emptyset)$ in place of $\delta(e)$. The ($\Longleftarrow$) side is also just as before, except that we employ Grove system $\Gamma = \{\delta(\emptyset), N\}$, and choose $+$ to agree with Γ when $\delta(\emptyset)$ is being updated. The rest of the argument is as before, with $\delta(\emptyset)$ replacing $\delta(e)$ everywhere.

Proof of 2: ($\Longrightarrow$) Suppose that δ is sequentially AGM with arbitrary revisions. Suppose $\delta(e^\frown q) = \delta(e) + ([e^\frown q] \cap \mathcal{A}) \neq \emptyset$. By (AGM 1), $\delta(e) + ([e^\frown q] \cap \mathcal{A}) \subseteq ([e^\frown q] \cap \mathcal{A})$, so $\delta(e^\frown q) \cap ([e^\frown q] \cap \mathcal{A}) \neq \emptyset$ and hence $\delta(e^\frown q) \cap [e^\frown q] \neq \emptyset$. The repetitive case is the same with $\delta(\emptyset)$ in place of $\delta(e)$.

($\Longleftarrow$) This side was already shown without appeal to quasi-consistency. □

In the proof, we showed that the extra set $\mathcal{A}$ chosen to revise on can always be $\delta(e^\frown q)$ itself. We also employed a trivial revision operator $+$ (i.e. the operator generated by full meet according to the Levi identity). Strengthening $+$ would endanger condition $(*)$ in the proof, possibly leading to a method whose conjectures are sometimes stronger than the corresponding conjectures of δ.

It is not intuitively clear to us why violations of internal stubbornness should be denounced as irrational, but it is at least straightforward (in the absence of computability considerations) to bring an arbitrary method into compliance with this principle without affecting its reliability or delaying its time of convergence. For suppose that method δ uniformly identifies the complete Λ truth, possibly violating internal stubbornness. Let γ agree with δ on data $\emptyset$. On data sequence $e^\frown q$, γ agrees with δ if this does not lead to a violation of internal stubbornness, and $\gamma(e^\frown q) = \delta(e^\frown q) - \gamma(e)$ otherwise. Then γ is internally stubborn in the sequential sense and γ uniformly identifies the complete Λ truth. The same modification works for a theory learner δ that piecemeal identifies the complete Λ truth.

7. MINIMAL REVISIONISM

There is an even weaker way for inductive methodology to interact with belief revision theory. According to this approach, revision theory is nothing but an explication of what it means to add a proposition to a theory that might be inconsistent with it, and an inductive method is allowed to retract on or to add whatever it pleases for whatever reason. Accordingly, δ is sequentially AGM in the *minimal sense* just in case there is some belief revision operator $+$ such that for each $e^\frown q$ there are $\mathcal{A}$ and $\mathcal{B}$ such that $\delta(e^\frown q)$ is the result of first contracting $\delta(e)$ by $\mathcal{A}$ and then revising the result by $\mathcal{B}$. The repetitive version of this concept is defined by replacing $\delta(e)$ with $\delta(\emptyset)$. On either interpretation, belief revision theory is entirely vacuous as a constraint on discovery methods.

Proposition 7 *Every method δ is a sequentially [repetitively]* AGM *method in the minimal sense.*

Proof: We begin with the sequential case. Let δ be given. Let $+$ be the operator represented by the trivial Grove systems defined in the proof of proposition 6. Let $-$ be defined according to the Harper identity as[11]

$$\mathcal{K} - \mathcal{A} = \mathcal{K} \cup (\mathcal{K} + \overline{\mathcal{A}}).$$

Then $\mathcal{K} - \mathcal{K} = \mathcal{K} \cup (\mathcal{K} + \overline{\mathcal{K}}) = \mathcal{K} \cup \overline{\mathcal{K}} = \mathcal{N}$ (the second to last identity is by $(*)$ in the proof of proposition 6.) Moreover, for each $\mathcal{A}, \mathcal{N} + \mathcal{A} = \mathcal{A}$. Hence, $\delta(e^\frown q) = (\delta(e) - \delta(e)) + \delta(e^\frown q)$. The repetitive case is similar. $\Box$

8. CONSERVATISM, RELIABILITY, AND LOGIC

We began with a fundamental distinction between conservative and reliabilist approaches to the problem of inductive inference. According to the former, our beliefs should be repaired in the most elegant possible manner when they are revised. According to the latter, the aim is to find the right answer whatever that answer might be. Belief revision theory belongs to the former perspective and formal learning theory belongs to the latter. In this paper, we have examined various different ways in which belief revision theory could be viewed

[11] In our propositional setting, the intersection of the deductive closures of two propositions $\mathcal{X}, \mathcal{Y}$ is just the union of all propositions including both $\mathcal{X}$ and $\mathcal{Y}$.

as as constraint on empirical inquiry, ranging from the view that inquiry is nothing but revision on the data to the proposal that revision is nothing but an explication of what it means to force a proposition into a system of beliefs, leaving inquiry to decide what to add or subtract. The result of our investigation is that the weaker interpretations impose few short-run restrictions on inquiry, whereas the stronger ones imply obstacles that must be carefully circumvented if inquiry is to be reliable.

This study illustrates two ways in which logic can be brought to bear on the problem of understanding inductive inquiry. One style of philosophical logic proposes intuitive normative principles and tests and refines these principles in a quasi-empirical way, by deducing particular consequences to compare against one's intuitions, by providing alternative representations of the principles, and by proving soundness and completeness theorems for the principles according to some illuminating semantic interpretation. The extensive work on belief revision theory fits into this approach. In this paper we have attempted to illustrate how the pertinent logical questions change when one shifts one's focus from coherence to reliability. Reliabilist applications of logic are more akin to the theory of computability than to usual work in philosophical logic. Whereas philosophical logicians tend to analyze systems reflecting norms derived from direct intuition or from particular case examples, the logic of reliability concerns the intrinsic solvability of inductive problems. The logical tools employed are programming, definability, and diagonalization rather than semantic interpretations, completeness and representations.

We hope to have illustrated how these two approaches can usefully interact. The logical analysis of intuitive axioms of rationality can motivate local side-constraints on inductive methods (e.g., internal stubbornness and timidity) that can enrich reliability analysis.[12] On the other side, we have shown in a preliminary way how reliability can be imposed as an extra axiom in a normative theory of belief revision, leading to enriched representations of acceptable inductive practice, as in the representation of truly reliable AGM methods by Grove systems of a special kind.

We hope that the range of questions raised by this preliminary study underscores the fruitfulness of reliability considerations in the study of belief revision. Is there an elegant representation of the reliable AGM methods that may produce infinitely many false conjectures? Are all problems solvable with finitely many false conjectures solvable by AGM methods? What happens to our results when more stringent constraints are imposed on revision than just the AGM axioms? What happens when we allow facts about the data stream to arrive in an arbitrary order?[13] And perhaps most importantly, how restrictive are the AGM principles for computable inquiry? Our non-restrictiveness results in this paper make use of highly idealized constructions in which un-

[12] It is interesting how naturally the Grove systems could be "programmed" as inductive methods.

[13] Cf. (Martin and Osherson, 1995).

computable logical relations must be decided. In light of the strong negative results concerning consistency in similar settings (Kelly and Schulte, 1995a; Kelly and Schulte, 1995b), one would expect the AGM principles to be restrictive for computable inquiry for that reason alone.

APPENDIX: PROOF OF PROPOSITION 2

Proposition 2 *The following statements are equivalent:*

1. *δ is data retentive, consistent, stubborn, and timid.*
2. *δ is sequentially* AGM.
3. *δ is repetitively* AGM.

Proof: To see that (3) implies (1), let δ be repetitively AGM. Then for some AGM operator $+$, we have that for each $e, \delta(e) = \delta(\emptyset) + [e]$. δ is data retentive by (AGM 1) and is consistent by (AGM 2). Now suppose that $\delta(e) \cap [e^\frown q] \neq \emptyset$. Then by axiom (AGM 4), $\delta(e^\frown q) = \delta(\emptyset) + [e^\frown q] = \delta(\emptyset) + ([e] \cap [e^\frown q]) = (\delta(\emptyset) + [e]) \cap [e^\frown q] = \delta(e) \cap [e^\frown q]$. So δ is stubborn and timid. An even simpler argument establishes that (2) implies (1).

To see that (1) implies (3), let δ be data retentive, consistent, etc. To show that δ is also repetitively AGM, one must show not only that δ behaves like an AGM operator repeatedly updating $\mathcal{K}$, but that δ's domain can be extended to all empirical propositions in a manner that satisfies the AGM axioms. It is useful to employ the Grove representation for this purpose.

We will construct a single Grove system Γ for $\delta(\emptyset)$ that behaves just like δ, from which it follows by proposition 1 that δ is repetitively AGM. Define

$$\mathcal{S}_0 = \delta(\emptyset);$$

$$\mathcal{S}_{n+1} = \mathcal{S}_n \cup \bigcup_{lh(e)=n+1} \delta(e);$$

$$\mathcal{S}_\omega = \mathcal{N};$$

$$\Gamma = \{\mathcal{S}_\alpha : \alpha \leq \omega\}.$$

It is easy to verify that Γ is a Grove system for $\delta(\emptyset)$.

Next, we show by induction on the length of e that $\Gamma([e]) = \delta(e)$. It then follows immediately by proposition 1 that δ is repetitively AGM. In the base case, we have: $\Gamma([\emptyset]) = \Gamma(\mathcal{N}) = \mathcal{S}_0 = \delta(\emptyset)$.

For the inductive case, suppose $\Gamma([e]) = \delta(e)$. Let q be given.

Case 1: suppose that $[e^\frown q] \cap \Gamma([e]) \neq \emptyset$. Then $\Gamma_{[e^\frown q]} = \Gamma_{[e]}$, so $\Gamma([e^\frown q]) = \Gamma([e]) \cap [e^\frown q] = \delta(e) \cap [e^\frown q]$ by the inductive hypothesis. But since δ is both timid and stubborn, the case hypothesis yields that $\delta(e^\frown q) = \delta(e) \cap [e^\frown q]$, and hence $\delta(e^\frown q) = \Gamma([e^\frown q])$.

Case 2: $[e^\frown q] \cap \Gamma([e]) = \emptyset$. Let $\mathrm{lh}(e) = n$. First we establish that

$$(*)\ \Gamma_{[e^\frown q]} = \mathcal{S}_{n+1}.$$

Since δ is consistent, $\delta(e^\frown q) \neq \emptyset$. Since δ is data retentive, $\delta(e^\frown q) \subseteq [e^\frown q]$. But $\delta(e^\frown q) \subseteq \mathcal{S}_{n+1}$, by definition of $\mathcal{S}_{n+1}$. Hence, $[e^\frown q] \cap \mathcal{S}_{n+1} \neq \emptyset$. Suppose for reductio that for some $i \leq n$ there is some $\varepsilon \in \mathcal{S}_i \cap [e^\frown q]$. Let k be the least such i. By choice of k and the definition of $\mathcal{S}_k$, there is an e' of length k such that $\varepsilon \in \delta(e')$. Since δ is data retentive, $\varepsilon \in [e']$. So $[\varepsilon|k], [\varepsilon|k+1], \ldots [\varepsilon|n], [\varepsilon|n+1]$ are all consistent with $\delta(\varepsilon|k)$. Since δ is both stubborn and timid, $\delta(e) = \delta(\varepsilon|n) = \delta(\varepsilon|k) \cap [\varepsilon|n]$. Hence, $[\varepsilon|n+1] \cap \delta(\varepsilon|n) \neq \emptyset$. But $\varepsilon|n = e, \varepsilon|n+1 = e^\frown q$ and by induction hypothesis, $\delta(e) = \Gamma([e])$, so $[e^\frown q] \cap \Gamma([e]) \neq \emptyset$, contrary to the case hypothesis.

Thus we have $(*)$. Hence, $\Gamma([e^\frown q]) = (\mathcal{S}_{n+1} - \mathcal{S}_n) \cap [e^\frown q]$. But by definition, $(\mathcal{S}_{n+1} - \mathcal{S}_n) = \bigcup_{lh(e)=n+1} \delta(e)$. Since δ is data retentive, $\delta(e^\frown q) \subseteq [e^\frown q]$. Let $e'' \neq e^\frown q$ be of length $n+1$. Then $[e''] \cap [e^\frown q] = \emptyset$.[14] Since δ is data retentive, $\delta(e'') \cap [e^\frown q] = \emptyset$. Hence, $\mathcal{S}_{n+1} \cap [e^\frown q] = \delta(e^\frown q)$.

Now we argue from (1) to (2). Let δ be data retentive, consistent, etc. Let proposition $\mathcal{K}$ be given. If $\mathcal{K}$ is not in the range of δ, then let $\Gamma_{\mathcal{K}}$ be an arbitrary Grove system. If $\mathcal{K}$ is in the range of δ then we define $\Gamma_{\mathcal{K}}$ as follows. First, define:

$D^0_{\mathcal{K}} = \{\emptyset\}$;

$D^{n+1}_{\mathcal{K}} = \{e^\frown q : \mathrm{lh}(e) = n \text{ and } \delta(e) = \mathcal{K}\}$.

Now define

$\mathcal{S}^0_{\mathcal{K}} = \mathcal{K}$;

$\mathcal{S}^{n+1}_{\mathcal{K}} = \mathcal{S}^n_{\mathcal{K}} \cup \bigcup_{e \in D^{n+1}_{\mathcal{K}}} \delta(e)$;

$\mathcal{S}^\omega_{\mathcal{K}} = \mathcal{N}$;

$\Gamma^{\mathcal{K}} = \{\mathcal{S}^\alpha_{\mathcal{K}} : \alpha \leq \omega\}$.

Since the $\delta(e)$'s added at each stage are mutually disjoint, one can show just as in the preceding argument that

$$\delta(\emptyset) = \Gamma^{\delta(\emptyset)}([\emptyset])$$

and

$$\delta(e^\frown q) = \Gamma^{\delta(e)}([e^\frown q]).$$

□

REFERENCES

Gärdenfors, P. (1988). *Knowledge In Flux: modeling the dynamics of epistemic states.* Cambridge: MIT Press.

[14] Our argument turns heavily on the fact that two distinct data sequences e, e' of the same length generate disjoint propositions $[e], [e']$. (Martin and Osherson, 1995, Prop. 21) contains a construction of a reliable repetitive AGM method that does not require this assumption, employing a well-ordering of theories.

Grove, A. (1988) "Two Modellings For Theory Change," *Journal of Philosophical Logic* 17: 157–170.

Kelly, K. (1995). *The Logic of Reliable Inquiry.* Oxford: Oxford University Press.

Kelly, K. and Glymour, C. (1990). "Theory Discovery from Data with Mixed Quantifiers," *Journal of Philosophical Logic* 19: 1–33.

Kelly, K. and Schulte, O. (1995a) "The Computable Testability of Theories Making Uncomputable Predictions," *Erkenntnis.* 43:29–66.

Kelly, K. and Schulte, O. (1995b) "Church's Thesis and Hume's Problem". *These proceedings.*

Martin, E. and Osherson, D. (1995). "Scientific discovery based on belief revision". *These proceedings.*

Osherson, D., Stob, M. and Weinstein, S (1986). *Systems That Learn.* Cambridge, Mass: MIT Press.

Popper, K. (1968). *The Logic Of Scientific Discovery.* New York: Harper.

Putnam, H. (1963). " 'Degree of Confirmation' and Inductive Logic," In *The Philosophy of Rudolph Carnap*, ed. A. Schilpp. La Salle, Ill.: Open Court.

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

Kevin Kelly and Oliver Schulte
Department of Philosophy
Carnegie Mellon University
Pittsburgh, PA 15213
USA
E-mail: kevin.kelly@andrew.cmu.edu, oliver.schulte@andrew.cmu.edu

Vincent Hendricks
Department of Philosophy
University of Copenhagen
Email: vincent@coco.ihi.ku.dk