Iterated Belief Revision,

Reliability

and

Inductive Amnesia

Kevin Kelly

February 5, 1998

Technical Report No. CMU-PHIL-88

Philosophy

Methodology

Logic

Carnegie Mellon

Pittsburgh, Pennsylvania 15213

Iterated Belief Revision, Reliability, and Inductive Amnesia

Kevin T. Kelly
Department of Philosophy
Carnegie Mellon University

February 27, 1998

0.1 Introduction

According to the familiar, Bayesian account of probabilistic updating, full beliefs change by *accretion*: in light of new information consistent with one's current beliefs, one's new belief state is the the result of simply adding the new information to one's current beliefs and closing under deductive consequence. Inductive generalizations that extend both one's current beliefs and the new information provided are not licensed, although the new information may increase the agent's *degree* of belief in such a proposition.[1] This account breaks down when new information contradicts the agent's current beliefs, for accretive updating leads, in this case, to a contadictory belief state from which further accretion can never escape. *Belief revision theory* aims to provide an account of how to update full belief so as to preserve consistency when one's current beliefs are refuted by the new information provided. Belief revision theory has attracted attention in a number of areas, including data base theory (Katsuno and Mendelson 1991), the theory of conditionals (Boutilier 93; Levi 96; Arlo-Costa 1997), the theory of causation (Spohn 1988, 1990; Goldszmidt and Pearl 94), and game theory (Samet 1996).

A *belief revision method* is a rule for modifying an agent's overall *epistemic state* in light of new information. An agent's beliefs are only part of her epistemic state, which also specifies an *implausibility* assignment over possibilities *inconsistent* with the agent's full beliefs. Upon receipt of new information, these degrees of implausibility determine the agent's revised belief state according to the following *belief revision rule*: the new belief state is the proposition satisfied exactly by the most plausible possibilities satisfying the newly received information.[2] According to this rule, the character of the revised belief state depends on the character of the agent's initial epistemic state. If all possible worlds are assigned implausibility degree 0, the agent starts out as a *tabula rasa* with vacuous beliefs and updates by mere accretion, without taking any inductive risks. At the opposite extreme, consider an agent whose initial epistemic state is maximally *refined*, in the sense that all possible worlds are assigned distinct degrees of implausibility. Such an agent starts out fully convinced of a complete theory and retains this conviction until the theory is refuted, at which point she replaces it with the complete theory of the most plausible world consistent with the new information. The new theory may differ radically from its predecessor. Described this way, belief revision sounds like a process of "eliminative" or "enumerative" induction, in which a "bold conjecture" is retained until it is refuted, after which it is replaced with the first alternative theory (in a subjective "plausibility ranking") that is consistent with the new information provided (Popper 68; Kemeny 53; Putnam 63; Gold 67; Earman 92). Between these two extremes are agents with moderately refined initial states whose inductive leaps from one theory to another are correspondingly weaker.

The belief revision literature has focused on the aim of minimizing damage to the agent's epistemic (or belief) state when new information contradicting the agent's beliefs is received. The similarity between belief revision and eliminative induction suggests a natural, alternative aim for belief revision: namely, to arrive at strong, true, empirical beliefs on the basis of increasing information. This aim is largely unexplored in the belief revision literature,[3] but it has long been the principal focus of *formal learning theory*, the study of processes of sequential belief update that are *reliable*, or guaranteed to stabilize to true, informative beliefs on increasing, true information. The purpose of this paper is to bring learning theoretic analysis to bear on a variety of methods of iterated belief revision proposed by Spohn (1988), Boutilier (1993), Nayak (1994), Goldszmidt and Pearl (1994), and Darwiche and Pearl (1997).[4] A very

---

[1] A broadly Bayesian perspective is not bound to identify inductive methodology with updating by conditionalization. Genuinely inductive expansions may be justified by decision-theoretic considerations based on epistemic utility (e.g., Levi 81).

[2] Degrees of implausibility are not degrees of belief in the usual, Bayesian sense, since degrees of belief are defined only over possibilities consistent with the agent's full beliefs, and therefore cannot guide belief-contravening updates according to the rule just described.

[3] It is raised, informally, in (Gärdenfors 1988).

[4] For earlier applications of learning theoretic analysis to belief revision theory, cf. (Martin and Osherson 1995, 1996) and (Kelly, Schulte and Hendricks 1996).

1

simple model of learning is employed, in which the successive propositions received by the agent are true reports of successive outcomes of some discrete, sequential experiment. An *inductive problem* specifies how the outcome sequence might possibly evolve in the unbounded future. The agent's task is to stabilize to complete, true beliefs about the outcome sequence, regardless of which sequence in the inductive problem is taken to be actual.

The investigation yields an interesting mixture of positive and negative results. Some of the methods are *empirically complete*, in the sense that for each solvable learning problem, there exists an initial epistemic state for which the method solves it. Others *restrict* reliability, in the strong sense that there are solvable learning problems that they cannot solve no matter how cleverly we adjust their initial epistemic states. All of the restrictive belief revision methods considered can have their initial epistemic states adjusted so that they remember the past, and nearly all of them can be adjusted to eventually predict the future. So such a method has the odd property that it can remember the past perfectly but then it cannot eventually predict the future and it can eventually predict the future, but then it forgets some of the past. I refer to this odd limitation as *inductive amnesia*. Inductive amnesia is the sort of thing we would like rules of rationality to protect us from rather than impose on us.[5] Avoiding it can therefore function as a nontrivial, motivated constraint on proposed methods and principles of belief revision.

Among the inductively amnestic belief revision methods, it is of interest to determine which are more restrictive than others. To answer these questions, I introduce a hierarchy of increasingly difficult inductive problems based on the number of applications of Nelson Goodman's (1983) "grue" operation, which reverses the binary outcomes in a data stream from a given point onward. For each of the methods considered, I determine the hardest problem in this *grue hierarchy* that the method can solve, obtaining, thereby, a reliability "fingerprint" of the method.

It might be expected that a global consideration such as eventually finding the truth would impose only the loosest short-run constraints on concrete methods of belief revision. However, sharp and unexpected recommendations are obtainable. For example, several proposed belief revision methods are equipped with a parameter α, which is the amount by which the updated implausibility of a possibility exceeds its prior implausibility when it is refuted. Lower values of α may be interpreted as more stringent notions of "minimal" change since they correspond, in a sense, to less distortion of the original epistemic state.[6] Two of these methods (Spohn 1988, 1990; Darwiche and Pearl 1997) turn out to fail by the second level of the grue hierarchy if $\alpha = 1$ but succeed over the entire, infinite grue hierarchy if α is incremented to 2. The difference between 1 and 2 is innocuous in light intuitive coherence and symmetry considerations, but it is pivotal for reliability. It will be argued, moreover, that the infinite leap in reliability at $\alpha = 2$ is not a technical aberration, but rather reflects a fundamental epistemological dilemma between memory and prediction faced by iterated belief revision methods of the sort under consideration.

The purpose of this paper is not to argue that reliability considerations always win when they conflict with coherence, symmetry, or minimality of belief change. As in every case of conflicting aims, a personal balance must be sought. But if the ultimate balance is subjective, structural conflicts between intuitive rationality considerations and reliability are not. The isolation and investigation of such conflicts is therefore a suitable aim for objective, epistemological analysis. The following results are, admittedly, preliminary and subject to generalization and refinement along a number of dimensions. Nonetheless, I hope they illustrate how reliability analyses might usefully and routinely be carried out for proposed theories of iterated belief revision.

---

[5] The fact that the methods can all be programmed to possess perfect memory blocks the response that inductive amnesia is a matter of resource bounds rather than a defect in the updating rules subject to it.

[6] α may also be viewed as an assessment of the quality or reliability of the input information. The results of this paper are less pertinent to that interpretation.

0.2 Ordinal Implausibility

Let W be a set of possible worlds.[7] The agent's *epistemic state* at a given time is modelled as an *implausibility assignment (IA)*, which is a (possibly partial) ordinal-valued function r defined on W.[8] Possibilities that are not even in the domain of r are "beyond possible consideration" in the strong sense that they will never be consistent with the agent's belief state, no matter what information the agent might encounter in the future. For a given world w, let $[w]_r$, $[w]_r^{\leq}$, and $[w]_r^{<}$ denote, respectively, the set of all worlds equally, no more, or less implausible than w.

A proposition is identified with the set of all possible worlds satisfying it. The full belief state of r is defined to be the proposition satisfied exactly by the possible worlds of implausibility zero.

$$b(r) = r^{-1}(0).$$

Define the minimum degree of implausibility of worlds in E as follows:

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

It will also be convenient to refer to the lowest degree of implausibility that is strictly greater than the implausibility of each world in E:

$$r_{\text{above}}(E) = \min\{\alpha : \forall w \in E \cap \text{dom}(r), r(w) < \alpha\}.$$

If $\alpha \leq \beta$ then $-\alpha + \beta$ denotes the unique γ such that $\alpha + \gamma = \beta$ (i.e., $-\alpha + \beta$ is the order type of the "tail" that remains when the initial segment α is "deleted" from β). $r(.|E)$ is an ordinal valued function with domain $\text{dom}(r) \cap E$ such that for each w in this domain:

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

Then $r_{\min}(A|E)$ and $r_{\text{above}}(A|E)$ may be defined as follows:

$$r_{\min}(A|E) = (r(.|E))_{\min}(A).$$

$$r_{\text{above}}(A|E) = (r(.|E))_{\text{above}}(A).$$

0.3 Some Implausibility Revision Methods

An *implausibility revision method* takes an IA r together with an input proposition E and returns an updated IA r'.

We will consider the following examples. Perhaps the most obvious idea is simply to eliminate refuted worlds from one's ranking and to lower all the other worlds, keeping intervals of relative implausibility fixed, until the most plausible world touches bottom. This is what Spohn (1988) refers to as the conditional implausibility ranking given the data.

Definition 1 (conditioning) $r *_C E = r(.|E)$.

Conditioning throws away refuted worlds, so it cannot recover when later data contradict earlier data. The remaining proposals boost the implausibility of refuted worlds rather than disposing with them altogether. An idea very similar to conditioning, mentioned by Goldszmidt and Pearl (1994), retains the refuted worlds but sends them all to a safe "point at infinity" never to be seen again unless past data are contradicted by future data.[9] It will prove interesting to analyze a generalization of this proposal in which all refuted worlds are assigned a fixed ordinal α.

---

[7]The approach adopted in this section follows Spohn (1988).

[8]It is not generally accepted that degrees of implausibility are well-ordered. This assumption will be dropped in section 0.8.

[9]Goldszmidt and Pearl introduce this proposal as an instance of the method $*_{J,\alpha}$ below, with $\alpha = \omega$.

Definition 2 (The "all to α" method)

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

Another proposal boosts all refuted worlds just above all the non-refuted worlds, maintaining intervals of implausibility among refuted worlds and among non-refuted worlds but not between the two classes.

Definition 3 (The lexicographic method)

$$(r *_L E)(w) = \begin{cases} r(w|E) & \textit{if } w \in \textit{ dom}(r) \cap E \\ r_{above}(E|E) + r(w|W - E) & \textit{if } w \in \textit{ dom}(r) - E \\ \uparrow & \textit{otherwise.} \end{cases}$$

A variant of this method was defined by Spohn (1988), who rejected it because it is irreversible, fails to commute (the resulting IA depends on the order in which the data arrive) and places extreme importance on the data (the refuted worlds are put above all the non-refuted worlds rather than being shuffled in). Against these considerations, S. M. Glaister (1997) has argued that a generalization of this rule due to Nayak (1994) is uniquely characterized by plausible symmetry conditions.

At the opposite extreme, consider the method that drops the lowest worlds consistent with the new information to the bottom level, and that rigidly elevates all other worlds by one step, keeping their relative positions to one another fixed.

Definition 4 (The "minimal" method)

$$(r *_M E)(w) = \begin{cases} 0 & \textit{if } w \in E \cap b(r(.|E)) \\ r(w) + 1 & \textit{if } w \in \textit{ dom}(r) - (E \cap b(r(.|E))) \\ \uparrow & \textit{otherwise.} \end{cases}$$

In a sense, this is the minimum alteration of the epistemic state consistent with the principle that one's new belief state be the set of all most entrenched possibilities consistent with the new information. Boutilier's "natural method" (1993) generalizes this method to apply to total pre-orders on worlds rather than IAs.[10] Spohn (1998) describes a method of this kind and rejects it. It doesn't fare better in terms of reversibility and commutativity and, in Spohn's opinion, places too little importance on the data, since the method can easily end up admitting possibilities excluded by the information received at the previous stage.

Spohn recommends, instead, the following sort of method. As usual, sort the worlds at each level into those that are refuted by the current evidence and those that are not. Lower both groups of worlds, preserving distances within the two groups, until the lowest worlds in each group are at the bottom level. Now raise all of the refuted worlds together so that the lowest refuted words end up at level α.[11] Spohn shows that this rule can be represented as updating a nonstandard probability measure by Jeffrey's rule, so long as there are but countably many possible worlds assigned to each degree of implausibility. It is also shown by Spohn to be both reversible and commutative (if α is understood to be an adjustable parameter). Nor is it as "extreme" as the preceding rules. But according to this rule, the implausibility of a refuted world may actually go down when it is refuted, since α may be below its current position and it might be the lowest refuted world at this stage of inquiry.

---

[10]Boutilier's method is considered in section 0.8 below. Boutilier also considers the problem of updating on conditionals, which is not addressed in this paper.

[11]This is actually a special case of Spohn's proposal. In general, Spohn's rule updates on a partition of possible worlds, with a separate α for each cell of the partition. It is assumed that one such α is zero. Here I present only the special case of binary partitions.

Definition 5 (Spohn's "Jeffrey conditioning" method)

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

Darwiche and Pearl (1997) propose an interesting modification of Spohn's Jeffrey conditioning method. Instead of dropping the refuted worlds to the bottom level before elevating them by α, Darwiche and Pearl propose lifting the refuted worlds by α from their current position, whatever that might be. The rule is presented with $\alpha = 1$, presumably to minimize alteration of the epistemic state, but the possibility of raising the value of α is left open.

Definition 6 (Darwiche and Pearl) *Let α be an ordinal.*

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

Proponents of different belief revision methods have in mind different conceptions of minimal change and different assessments of the relative importance of minimality as opposed to other symmetry conditions. Such debates may be irresolvable. My purpose is to shift the focus of such debates to the relative abilities of the various methods to generate true, informative beliefs; a natural goal that distinguishes sharply and objectively between the above proposals.

0.4   Iterated Implausibility Revision as Inductive Inquiry

Iterated belief revision involves successive modifications of one's epistemic state as successive input propositions are received. Iteration of an implausibility operator over a sequence of propositions is defined recursively as follows:

1. $r * () = r$

2. $r * (E_0, \ldots, E_n, E_{n+1}) = (r * (E_0, \ldots, E_n)) * E_{n+1}.$

A belief revision agent starts out with an initial epistemic state r and sequentially updates her beliefs using a belief revision operator $*$, so we may identify the agent with a a pair $(r, *)$, which I refer to as an *implementation* of $*$. Such an agent determines a unique map from finite sequences of input propositions to new belief states as follows:

$$(r, *)(E_0, \ldots, E_n) = b(r * (E_0, \ldots, E_n)).$$

In general, an *inductive method* is a rule that produces an empirical hypothesis in response to a finite sequence of input propositions:

$$f(E_0, \ldots, E_n) = B_{n+1}.$$

Inductive methods are the usual objects of learning theoretic analysis. Since an implementation $(r, *)$ of a belief revision operator $*$ is a special kind of inductive method, it is directly subject to learning theoretic analysis.

0.4.1 Data Streams

Suppose a scientist who uses an inductive method f is faced with the task of studying the successive outcomes of experiments on some unknown system. We will suppose that the outcomes are discretely recognizable, and hence may be encoded by natural numbers.

The *data stream* generated by the system under study is just an infinite tape on which the code numbers of the successive outcomes of the experiment are written. The first datum arrives at stage 0, so a data stream is a total function e defined on the natural numbers. Let U denote the set of all data streams. An *empirical proposition* is a subset of U. In other words, the truth of an empirical proposition supervenes on the actual data stream.

Consider the scientist's idealized situation at stage n of inquiry. At that stage, she observes that the outcome for stage n is $e(n)$. Epistemically, she updates on the empirical proposition $[n, e(n)]$, which is defined to be the set of all data streams e' such that $e'(n) = e(n)$. The *initial segment* of the data stream scanned by stage n is

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

The *length* of of this sequence is defined to be n:

$$\mathrm{lh}(e(0), \ldots, e(n-1)) = n.$$

The *tail* of the data stream remaining to be scanned from stage n is:

$$n|e = (e(n), e(n+1), \ldots).$$

Prior to stage n, the scientist updates on the sequence of empirical propositions

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

Then her inductive method's output prior to observing $e(n)$ is just

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

Note that $[[e|n]]$ is not the same thing as the empirical proposition

$$[e|n] = \{e' \in U : e|n \text{ is extended by } e'\},$$

which states that the finite outcome sequence $e|n$ has occured. Rather, $[e|n]$ is the intersection of all the propositions $[i, e(i)]$ occurring in $[[e|n]]$. Now that these distinctions are clear, I will simplify notation by writing

$$f(e|n) = f([[e|n]]).$$

0.4.2 Empirical Questions

Inquiry has two cognitive aims, seeking truth and avoiding error.[12] Seeking truth involves relief from ignorance. One simple way to specify nontrivial content is to partition possibilities and to require that the outputs of the method eventually entail the true cell of this "target" partition. We may think of the partition as an *empirical question* and the cells of the partition as the potential *answers* to the question. Θ_0 denotes the singleton partition $\{\{e\} : e \in U\}$, which corresponds to the hardest empirical question "what is the complete empirical truth?" and Θ_1 denotes the trivial question $\{U\}$, answered by vacuously true beliefs.

---

[12]William James (1948), (Levi 1981).

6

0.4.3 Reliability in the Limit

Given an empirical question Θ, one may hope that one's method is guaranteed to halt with a correct answer to Θ. But no bell rings when science has found the truth,[13] suggesting the weaker requirement that inquiry eventually stabilize to a correct answer to Θ, perhaps without ever knowing when it has done so. Then we say that the method *identifies* an answer to Θ on e, or that the method *identifies* Θ on e for short.

It is not enough that a method *happen* to stabilize to the right answer in the actual world: scientific success should be more than opinionated luck. *Reliability* demands that a method succeed over some broad range K of possible data streams. One may think of K as the domain of the agent's initial epistemic state (i.e., the set of worlds that the agent might possibly admit as serious possibilities in the future). But one might also think of K simply as a range of possibilities over which the method can be shown to succeed, so that the method is more reliable insofar as K is larger (weaker). When the method identifies Θ on every data stream in K, we say that it identifies Θ given K. In the special case when the target partition is Θ_0, we may speak simply of identifying K.

Definition 7

    1. f identifies Θ given K just in case for each $e \in K$, for all but finitely many n, $e \in f(e|n) \subseteq \Theta(e)$.

    2. f identifies K just in case f identifies Θ_0 given K.

Identification requires that inquiry eventually arrive at complete, true beliefs both about the future and the past. One may weaken this requirement by countenancing incorrect or incomplete memories of the past, so long as these do not compromise predictive power. Then it will be said that method *projects* the complete future.

Definition 8 *f projects K just in case for each e in K and for all but finitely many $n, \emptyset \neq f(e|n) \subseteq [n|e]$.*

If projection looks forward, we may also look backward and ask if the method's conjecture at each stage consistently entails the data received thus far.

Definition 9 *f remembers K just in case for each e in K, for each $n, \emptyset \neq f(e|n) \subseteq [e|n]$.*

Clearly, f identifies K just in case f remembers K and f projects K. Intuitively, it seems as though perfect memory would only make reliable projection of the future easier. But for some of the methods introduced above, this is not true, as will be apparent shortly.

0.4.4 Identifiability, Restrictiveness and Completeness

Let M be the set of all inductive methods and let $M' \subseteq M$. Think of M' as a a proposed *architecture* or restriction on admissible inductive methods. For example, M' may reflect someone's "intuitive" ideas about rationality (e.g., that $f = (r, *)$, for some choice of $r, *$). Then we may say that Θ is *identifiable* by M' given K just in case there is an $f \in M'$ such that f identifies Θ given K, and similarly for the identifiability or projectability of K by M'. When $M' = M$, the explicit reference to M will be dropped.

Architecture M' is inductively *complete* just in case each identifiable Θ is identified by some method in M'. Otherwise, M' is inductively *restrictive*, in the sense that it prevents us from solving inductive problems we could have solved by other means.[14] In a similar manner, we may speak of completeness and restrictiveness with respect to function identification, projection, or memory. Restrictiveness raises

---

[13] This charming phrase is from William James (1948).
[14] The term "restrictiveness" is due to Osherson et al. (1986).

serious questions about the normative standing of a proposed account of rational inquiry, since it seems that rationality ought to augment rather than inhibit the search for truth.[15]

The main question before us is whether insistence on a particular belief revision operator * is restrictive (i.e., prevents us from answering inductive questions we could have answered otherwise). Let M^* denote the set of all inductive methods that implement the plausibility revision operator * (i.e., $M^* = \{(r, *) : r \in IA\}$. I say that * is complete or restrictive (in any of the above senses) just in case M^* is.

Some of the belief revision operators introduced above are restrictive. But their restrictiveness is manifested in a curious way: they are complete with respect to projection and they are complete with respect to memory, but they are restrictive with respect to identification. In other words, such methods can be implemented to remember or to project the future, but cannot be implemented to do both. Such a method is said to suffer from *inductive amnesia*. Inductive amnesia says that those who don't want to repeat history *should* forget it!

Since restrictiveness is a matter of preventing the solution of solvable problems, it is useful to characterize the set of solvable problems. Identifiability has an elegant topological characterization. Let K be a collection of data streams. Recall that for finite sequene ϵ, $[\epsilon] = \{e \in U : \epsilon$ is extended by $e\}$. A K-fan is a proposition of form $[\epsilon] \cap K$. Then we say S is K-*open* (or open in K) just in case S a union of K-fans. S is K-*closed* just in case $K - S$ is K-open.

Proposition 1 (characterization theorem for partition identification) *Let $\Theta[K]$ denote the restriction of Θ to K (i.e, $\{C \cap K : C \in \Theta\}$). Then Θ is identifiable given K just in case $\Theta[K]$ is countable and each cell in $\Theta[K]$ is a countable union of K- closed sets.*[16]

Proof: (Kelly 96). ⊣
    The characterization of function identifiability is even simpler:

Proposition 2 (characterization theorem for identification) *The following propositions are equivalent:*
1. K is identifiable; 2. K is projectable; 3. K is countable.

Proof: In Appendix I. ⊣
    Projectability and identifiability are equivalent with respect to the collection of all possible inductive methods, but not when we restrict attention to methods implementing an inductively amnestic revision operator *.

0.4.5 Counting Retractions

No scientist likes to retract. The social stigma associated with retraction reflects the painful choices and costly conceptual retooling that scientific revolutions entail (Kuhn 1970). Counting retractions provides a refinement of identification results. If Θ is identifiable, we can ask how many retractions are required in the worst case to identify Θ and demand that a method never exceed this bound.

Definition 10

1. $retractions(f, e) = |\{k : f(e|k) \not\subseteq f(e|k+1)\}|$.

2. f *identifies* K *with n retractions just in case f identifies K and for each e in K, $retractions(f, e) \leq n$.*

3. K *is identifiable with n retractions just in case f identifies K and for each e in K, $retractions(f, e) \leq n$.*

Identification with n retractions has a natural characterization in terms of Spohn's implausibility assignments independently of any choice of operator, a pleasant and revealing connection between learning theory and belief revision.

Proposition 3 (characterization of n retraction identifiability) Θ *is identifiable given K with at most n retractions just in case there is an r such that $rng(r) = \{0, \ldots, n\}$, $K \subseteq dom(r)$ and for each cell $C \in \Theta$, for each $k \leq n$,*

1. *C is open (and hence clopen) in $r^{-1}(k)$ and*

2. *$\bigcup_{i=1}^{k} r^{-1}(i)$ is closed in $dom(r)$.*

Proof: In Appendix I. $\dashv$

e is *isolated* in $S \subseteq U$ just in case for some n, $[e|n] \cap S \subset \{e\}$ (i.e., $\{e\}$ is clopen in S).

Proposition 4 (characterization of n retraction function identifiability) K *is identifiable with n retractions just in case there is an r such that $rng(r) = \{0, \ldots, n\}$, $K \subseteq dom(r)$ and for each $e \in K$, e is isolated in $[e]_{\bar{r}}^{\leq}$.*

Proof: In Appendix I. $\dashv$

0.5 Some Diachronic Properties of Implausibility Revision

Three diachronic properties of implausibility revision operators have particular relevance for reliability considerations. The first requires that the operator always produce new beliefs consistent with the current datum and the domain of the current IA. All belief revision theorists insist on this requirement and all the methods under consideration satisfy it.

Definition 11 (local consistency) $(r, *)$ *is locally consistent just in case for all $(A_1, \ldots, A_{n+1})$ such that $dom(r * (A_1, \ldots, A_n)) \cap A_{n+1} \neq \emptyset$,*

$$A_{n+1} \cap b(r * (A_1, \ldots, A_{n+1})) \neq \emptyset.$$

The next property requires preservation of the implausibility ordering among worlds satisfying all the input propositions received so far. This does not entail that the ordinal distances between such possibilities are preserved (gaps may appear or disappear).

Definition 12 (positive order-invariance) $(r, *)$ *is positively order-invariant just in case for all $(A_1, \ldots, A_n)$ such that $n > 0$, for all $w, w' \in dom(r) \cap A_1 \cap \ldots \cap A_n$,*

1. *$w, w' \in dom(r * (A_1, \ldots, A_n))$ and*

2. *$r(w) \leq r(w') \Leftrightarrow (r * (A_1, \ldots, A_n))(w) \leq (r * (A_1, \ldots, A_n))(w')$.*

A stricter property requires, as well, preservation of the ordinal distances among worlds consistent with all the data received so far.

Definition 13 (positive invariance) $(r, *)$ *is positively invariant just in case for all $A_1, \ldots, A_n)$ such that $n > 0$, for all $w, w' \in dom(r) \cap A_1 \cap \ldots \cap A_n$,*

1. *$w, w' \in dom(r * (A_1, \ldots, A_n))$ and*

2. *$r(w) - r(w') = (r * (A_1, \ldots, A_n))(w) - (r * (A_1, \ldots, A_n))(w')$.*

Local consistency and positive order-invariance say nothing about what to do with worlds that do not satisfy E. One requirement, reflecting high respect for the data, demands that each world satisfying E be strictly more plausible than every world failing to satisfy E. This property goes much farther than the requirement that the updated belief set $b(r * E)$ entail E. It governs the overall implausibility structure concerning even remotely plausible worlds.

Definition 14 (positive precedence) $(r, *)$ *is positively precedent just in case for all* $(A_1, \ldots, A_n)$, *for all* $w \in dom(r) \cap A_1 \cap \ldots \cap A_n$, *for all* $w' \notin dom(r) \cap A_1 \cap \ldots \cap A_n$,

1. $w' \in dom(r * (A_1, \ldots, A_n))$ *and* $w' \notin dom(r * (A_1, \ldots, A_n))$ *or*

2. $w, w' \in dom(r * (A_1, \ldots, A_n))$ *and* $(r * (A_1, \ldots, A_n))(w') > (r * (A_1, \ldots, A_n))(w)$.

For each of the properties just defined, we say that $*$ has the property just in case $(r, *)$ has the property, for each IA r.

Local consistency, positive order-invariance and positive precedence are logically independent. Together, they force a belief revision method to behave in a manner that makes a great deal of sense if finding the truth is the goal of inquiry. Consider a method with all three properties. It starts out with a fixed implausibility ranking r on worlds. Upon updating on E, positive precedence requires that all the non-E worlds are either weeded out altogether (they are not even in the domain of $(r * E)$ or are sent to a "safe" place beyond all the E worlds). By positive order-invariance, the E worlds remain ranked as they were before, (the ordinal intervals between two E-worlds may stretch or contract, however). By local consistency, the lowest of these E-worlds must drop to the bottom of the revised IA. As inquiry proceeds, such a method continues to weed out the non-E worlds and to conjecture the most plausible remaining worlds, according to a fixed implausibility ranking, so eventually the actual world migrates to the bottom of the ranking and the method's belief state is true forever after.

This kind of procedure has long been entertained under a variety of headings. In the philosophy of science it has been referred to as the method of bold conjectures and refutations (Popper 1968) or the hypothetico-deductive method (Kemeny 1953, Putnam 1963). In the learning theoretic literature it is referred to as the enumeration method (Gold 1967). Such a method is natural if one's aim is to converge to true, sufficiently strong beliefs. I refer to an operator that satisfies all three properties as an *enumerate and test* operator. Then we have:

Proposition 5 *If* Θ *is identifiable given* K *then there exists an* r *such that*

1. $rng(r) \subseteq \omega$ *and*

2. *for each* $*$ *such that* $(r, *)$ *enumerates and tests,* $(r, *)$ *identifies* Θ *given* K.

Proof: In Appendix II. ⊣

Now suppose that $(r, *)$ is locally consistent and positively order-invariant but does not satisfy positive precedence. Then the method still maintains a fixed ranking of implausibility over the E worlds, but non-E worlds may fail to rise above all the E worlds. Hence, it is possible for them to return, eventually, to the bottom of the ranking as inquiry continues. From the viewpoint of inductive methodology, this means that a refuted world may reappear to remain forever in the agent's beliefs, even if all the data received are mutually consistent. In other words, the agent may irrevocably forget the past. It is not difficult to choose particular initial epistemic states that lead such a method to forget. Inductive amnesia is the much less trivial situation in which *every* initial epistemic state that ensures that the method reliably predicts the future also causes it to forget some past datum.

Although $*_{DP,n}$ does not satisfy positive precedence, it satisfies the following, weakened version of positive precedence. The operator $*_{S,n}$ lacks this property because possibilities may backslide upon refutation.

Proposition 6 (climbing lemma) *Suppose $r(e), r(e'), n$ are finite and $r(e) > r(e')$. Then*

$$(r *_{DP,n} [[e|k]])(e) - (r *_{DP,n} [[e|k]])(e') \leq (r(e) - r(e')) - n\rho_k(e, e').$$

Proof: By induction on $\rho_k(e, e')$. ⊣

The following properties, like local consistency, are axioms of the AGM theory of belief revision (Gärdenfors 88) and are satisfied by all the belief revision operators under consideration.

Definition 15 (timidity and stubbornness) $(r, *)$ *is timid [stubborn] just in case for each* $(A_1, \ldots, A_{n+m})$ *such that* $(A_{n+1} \cap \ldots \cap A_{n+m}) \cap b(r * (A_1, \ldots, A_n)) \neq \emptyset$,
$b(r * (A_1, \ldots, A_n)) \cap (A_{n+1} \cap \ldots \cap A_{n+m}) \subseteq [\supseteq] b(r * (A_1, \ldots, A_{n+m}))$.

Timidity and stubbornness force full belief to evolve by mere accretion (according to the standard Bayesian approach) until one's full beliefs are refuted by new information. All enumerate-and-test methods are timid and stubborn and enumerate and test methods are all complete inductive architectures (proposition 5), which provides something of a reliabilist motivation for these properties.[17] Once positive precedence is dropped, however, timidity and stubbornness assume a more sinister aspect, since these properties are involved in each of the negative arguments presented in this paper.

Proposition 7 *The following table specifies which of the above properties hold of the operators under consideration regardless of the choice of r and of α.*

diachronic properties	C	L	DP, α	S, α	A, α	M
pos. order-invariance	yes	yes	yes	yes	yes	yes
pos. invariance	yes	yes	yes	yes	yes	no
local consistency	yes	yes	yes	yes	yes	yes
positive precedence	yes	yes	no	no	no	no
timidity	yes	yes	yes	yes	yes	yes
stubbornness	yes	yes	yes	yes	yes	yes

The following table summarizes the changes in the above table when it is assumed that $\alpha \geq r_{above}(dom(r))$.

diachronic properties	C	L	DP, α	S, α	A, α	M
positive precedence	yes	yes	yes	yes	yes	no

Proof: Induction on the stage of inquiry and some simple examples. ⊣

0.5.1 Inductive Completeness Theorems

In light of the preceding discussion of enumerate-and-test methods, it should come as little surprise that methods like $*_C, *_L, *_{DP,\omega}, *_{S,\omega}$, and $*_{A,\omega}$ are complete strategies for partition identification.

Proposition 8 (complete partition identification methods) *The following statements are all equivalent:*

1. Θ is identifiable given K;

2. Θ is identifiable by $*_C$ given K;

3. Θ is identifiable by $*_L$ given K;

---

[17]Positive invariance keeps unrefuted worlds at the bottom of the ranking below all other non-refuted worlds. Positive precedence sends all refuted worlds permanently above the non-refuted worlds. And local consistency ensures that the lowest of the non-refuted worlds stay down, so we have timidity and stubbornness.

4. Θ *is identifiable by* $*_{DP,\omega}$ *given* K;

5. Θ *is identifiable by* $*_{S,\omega}$ *given* K;

6. Θ *is identifiable by* $*_{A,\omega}$ *given* K.

Hence, $*_C, *_L, *_B, *_{DP,\omega}, *_{S,\omega}, *_{A,\omega}$ *are complete inductive architectures for partition identification.*

Proof: In Appendix II. ⊣

The following result concerns methods that are complete architectures for identification with n retractions. Recall that problems solvable with n retractions can be packed into an initial epistemic state whose highest level is n (proposition 3). Methods $*_{A,n+1}, *_C, *_L, *_{DP,n+1}$, and $*_{S,n+1}$ safely launch refuted worlds above all non-refuted worlds in such an ordering. Since the truth drops one level at each retraction, convergence occurs by the nth retraction.

Proposition 9 (n retraction completeness for partitions) *If* Θ *is identifiable given* K *with at most* n *retractions, then* $*_C, *_L, *_{DP,n+1}, *_{S,n+1}, *_{A,n+1}$ *can identify* Θ *given* K *with at most* n *retractions.*

Proof: In Appendix II. ⊣

The following completeness results concern the narrower problem of function identification. It is left open whether item 7 can be extended to arbitrary partition identification.

Proposition 10 (complete function identification methods) *The following propositions are equivalent:*

1. K is identifiable;

2. K is projectable;

3. K is identifiable by $*_C$;

4. K is identifiable by $*_L$;

5. K is identifiable by $*_{S,\omega}$;

6. K is identifiable by $*_{A,\omega}$;

7. K is identifiable by $*_{DP,2}$;

8. K is projectable by $*_{DP,1}$;

9. K is projectable by $*_{S,1}$.

Proof: In Appendix II. ⊣

Most of these equivalences follow from the preceding proposition and concern methods that boost refuted possibilities above all "live" possibilities. A surprising exception is the fact that $*_{DP,2}$ is a complete function identification architecture.[18] To prove completeness, one must construct, for each K, an epistemic state r such that $(r, *_{DP,2})$ identifies K. Here is how this can be done in the special case in which all elements of K are finite variants of one another. First, we define the *difference set* of all positions on which two data streams differ:

$$\Delta(e, e') = \{i \in \omega : e(i) \neq e'(i)\}.$$

Then define Hamming distance to be the size of the difference set.

$$\rho(e, e') = |\Delta(e, e')|.$$

---

[18] It is left open whether this can be extended to the case of partition identification.

Hamming distance is readily seen to be an extended metric on U. Given a fixed data stream e_0 we can construct an epistemic state

$$r_{e_0}^H(e) = \rho(e_0, e)$$

on K. This "Hamming" state has the nice property that a data stream e' that is n steps below the true data stream e differs from e in at least n positions. When $\alpha = 2$, e' moves up with respect to e at least two steps each time one of these n positions is seen, so e' ends up at least one step above e after all of these positions have been observed. The full completeness theorem is proved by a similar but somewhat more complicated construction.

0.6 The Grue Hierarchy

To show that a methodological recommendation restricts reliability, one must find an otherwise easily solvable problem that the recommended method fails to solve, no matter how its initial epistemic state is arranged. This end is served admirably by an unfamiliar application of a familiar idea due to the philosopher Nelson Goodman (1983).[19] Let 0 represent a "green" outcome and let 1 represent a "blue" outcome. Then a "grue_n" outcome is either a green outcome by stage n or a blue outcome after stage n. The everywhere green data stream is the everywhere 0 sequence and the everywhere grue_n sequence is a sequence of n 0s followed by all 1s. More generally, let $\neg b$ denote the Boolean complement of b. Let B denote the set of all Boolean-valued data streams. Then if $e \in B$, let $\neg e$ denote the outcome stream in which each outcome occurring in e is reversed (i.e., $(\neg e)(n) = \neg e(n)$). Now define the *grue* operation as follows:

$$e \ddagger n = (e|n)\neg(n|e).$$

In other words, $(e \ddagger n)(i) = e(i)$ if $i < n$ and $= \neg e(i)$ otherwise.

Grue operations are commutative:[20]

$$(e \ddagger n) \ddagger m = (e \ddagger m) \ddagger n.$$

Also, gruing twice in the same place yields the original data stream. Hence, each composed grue operation can be represented by the set S of positions that have grue operations applied an odd number of times. Let $e \ddagger S$ denote the (unique) data stream that results from applying, in any order, any odd number of grue operations at positions in S and any even number of grue operations (possibly zero) at all other positions.

Now given $K \subseteq B$, we can define a hierarchy of ever more complex inductive problems as follows:

Definition 16 (The Grue Hierarchy) *Let* $K \subseteq B$.

1. $g^n(K) = \{e \ddagger S : |S| = n \text{ and } e \in K\}$.

2. $G^n(K) = \bigcup_{i \le n} g^n(K)$.

3. $G^\omega(K) = \bigcup_{i < \omega} g^n(K)$.

The even grue hierarchy $G_{\text{even}}^n(K)$ is defined similarly, except that (1) is replaced with:

$$g_{\text{even}}^n(K) = \{e \ddagger S : |S| = 2n \text{ and } e \in K\}.$$

The distinction between the even and the full grue hierarchies is important for some of the belief revision methods under consideration, but it makes no difference to identifiability when no extra constraints are imposed on the scientist's inductive method:

---

[19] Goodman was not interested in constructing unsolvable inductive problems. His purpose was to show that "constancy" of the data stream is not preserved under translation, and hence is not a logical basis for a theory of confirmation.

[20] Let $\mathcal{G}$ denote the closure under composition of the set of grue all functions: $\{(. \ddagger k) : k \in \omega\}$. $(\mathcal{G}, \circ)$ is an Abelian group in which each element is its own inverse.

Proposition 11 *For all $n \in \omega, e_0 \in B$,*

    1. $G^n(e_0), G^n_{even}(e_0)$ are identifiable with n retractions but not with $n-1$ retractions.

    2. $G^\omega(e_0), G^\omega_{even}(e_0)$ are identifiable but not under any fixed bound on the number of retractions performed.

Proof: In Appendix I. ⊣

0.7   The Main Result

The following proposition determines exactly which problems in the grue hierarchy and the even grue hierarchy each of the methods under consideration can solve.

Proposition 12 (The grue scale) *For each $e_0 \in B$, the following table specifies which problems in the grue and even grue hierarchies each of the methods under consideration can identify. The classifications are optimal, in the sense that no lower value of α than the one reported suffices for identification of the corresponding problem.*

problem	M	A, α	S, α	DP, α	L	C
$G^\omega(e_0)$	no	$\alpha = \omega$	$\alpha = 2$	$\alpha = 2$	yes	yes
⋮	⋮	⋮	⋮	⋮	⋮	⋮
$G^n(e_0)$	no	$\alpha = n+1$	$\alpha = 2$	$\alpha = 2$	yes	yes
⋮	⋮	⋮	⋮	⋮	⋮	⋮
$G^2(e_0)$	no	$\alpha = 3$	$\alpha = 2$	$\alpha = 2$	yes	yes
$G^1(e_0)$	no	$\alpha = 2$	$\alpha = 2$	$\alpha = 1$	yes	yes
$G^0(e_0)$	yes	$\alpha = 0$	$\alpha = 0$	$\alpha = 0$	yes	yes
$G^\omega_{even}(e_0)$	no	$\alpha = \omega$	$\alpha = 1$	$\alpha = 1$	yes	yes
⋮	⋮	⋮	⋮	⋮	⋮	⋮
$G^n_{even}(e_0)$	no	$\alpha = n+1$	$\alpha = 1$	$\alpha = 1$	yes	yes
⋮	⋮	⋮	⋮	⋮	⋮	⋮
$G^2_{even}(e_0)$	no	$\alpha = 3$	$\alpha = 1$	$\alpha = 1$	yes	yes
$G^1_{even}(e_0)$	no	$\alpha = 2$	$\alpha = 1$	$\alpha = 1$	yes	yes
$G^0_{even}(e_0)$	yes	$\alpha = 0$	$\alpha = 0$	$\alpha = 0$	yes	yes

Proof: Propositions 10 and 18 in Appendix II, proposition 21 in Appendix III, and propositions 28, 29, 30, and 31 in Appendix IV. ⊣

Most of the positive results in the table follow from more general completeness results already discussed. Noteworthy exceptions are the abilities of $*_{S,1}$ and $*_{DP,1}$ to identify $G^\omega_{even}(e_0)$. This contrasts markedly with the situation in the full grue hierarchy, in which these methods all fail by level three. The situation is quite different for $*_M$ and $*_{A,n}$, which see no improvement in the even grue hierarchy (proposition 29.2). The problem $G^\omega_{even}(e_0)$ is just the set of all finite variants of a given data stream e_0. The evolution of $*_{S,1}$ and of $*_{DP,1}$ in this problem can be pictured as follows. Suppose the method starts out with an initial epistemic state ranking each data stream according to its Hamming distance from a given data stream e_0. Suppose e is the truth. Then the set of data streams differing from e_0 only where e_o differs from e may be viewed as a finite dimensional hyper-cube whose dimensionality matches the total number of differences between e_0 and e. Think of this hypercube as resting balanced on the vertex labelled with e_0. To find the initial implausibility of a vertex, find the shortest path from the bottom of

the cube to that vertex. It is shown (proposition 31) that if we resrict attention to these possibilities, the sequential operation of both $*_{S,1}$ and $*_{DP,1}$ can be viewed as the rigid rotation of the hypercube from one vertex to another on a direct path to the true (originally uppermost) vertex e. After e is rotated to the bottom, the rotation stops and the method has converged to the truth!

The happy image of learning as hypercube rotation cannot be extended to the full grue hierarchy. Indeed, no matter where we insert the data stream $\neg e_0$ into the restriction of the Hamming ranking to $G^2_{\text{even}}(e_0)$, $*_{S,\alpha}$ fails even when $\alpha = 2$.

Proposition 13 *For each $r \supseteq r^H_{e_0}|G^2_{\text{even}}$, $(r, *_{S,2})$ does not identify $G^2_{\text{even}}(e_0) \cup \{\neg e_0\}$.*

Proof: Appendix III. ⊣

Why should adding a single possibility to $G^\omega_{\text{ven}}(e_0)$ matter so much? Consider the case in which $\alpha = 1$. If we place $\neg e_0$ at an infinite level of the Hamming ranking, then it never falls the infinite distance to the bottom of the ranking, since the infinitely many elements of $G^1(e_0)$ occurring lower than $\neg e_0$ are never all refuted. If we insert $\neg e_0$ at a finite level, then $\neg e_0$ is below some other data stream e agreeing with $\neg e_0$ as far as we please. If e happens to be the truth, then the cube rotates as usual until $\neg e_0$ ends up at the bottom vertex. By timidity and stubbornness, $\neg e_0$ occupies the bottom level of the epistemic state until its first position differing from e is observed. By positive invariance, the presence of $\neg e_0$ at the bottom prevents the next vertex of the hypercube from dropping to the bottom level when the vertex currently at the bottom is refuted. Now instead of resting on a new vertex, the cube rotates up with one edge parallel to but not touching the plane. This causes e to become forever confused with a unit variant that was previously one level below it, as a simple simulation on a three-dimensional cube will illustrate.

By proposition 13, the initial state $*_{S,2}$ employs to identify $G^\omega(e_0)$ cannot be an extension of the Hamming ranking over $G^\omega_{\text{ven}}(e_0)$. Instead I employ a ranking based on *grue distance*, or the number of grue operations required to transform one data stream into another (cf. Appendix III).[21] Define the *grue set* for two data streams as follows:

$$\Gamma(e, e') = \{n \in \omega : [n = 0 \wedge e(n) \neq e'(n)] \vee [n > 0 \wedge \neg(e(n-1) = e'(n-1) \Leftrightarrow e(n) = e'(n))]\}.$$

The terminology is justified by the following fact.

Proposition 14 *$\Gamma(e, e')$ is the least $S \subseteq \omega$ such that e' can be obtained from e by applying grue operations only at positions in S.*

Proof: Omitted. ⊣

Now define the *grue distance* on B as follows:

$$\gamma(e, e') = |\Gamma(e, e')|.$$

In light of the preceding proposition, grue distance is the least number of grue operations required to transform e into e'. It is readily verified that grue distance is an extended metric over B. The initial epistemic state on B induced by grue distance is just:

$$r^G_{e_0} = \gamma(e_0, e').$$

For an algebraic perspective on the relationship between $r^G_{e_0}$ and $r^H_{e_0}$, define the Hamming and Goodman orders as follows (figure 1):

1. $e' \leq^{e_0}_G e'' \Leftrightarrow \Gamma(e_0, e') \subseteq \Gamma(e_0, e'')$.

2. $e' \leq^{e_0}_H e'' \Leftrightarrow \Delta(e_0, e') \subseteq \Delta(e_0, e'')$.
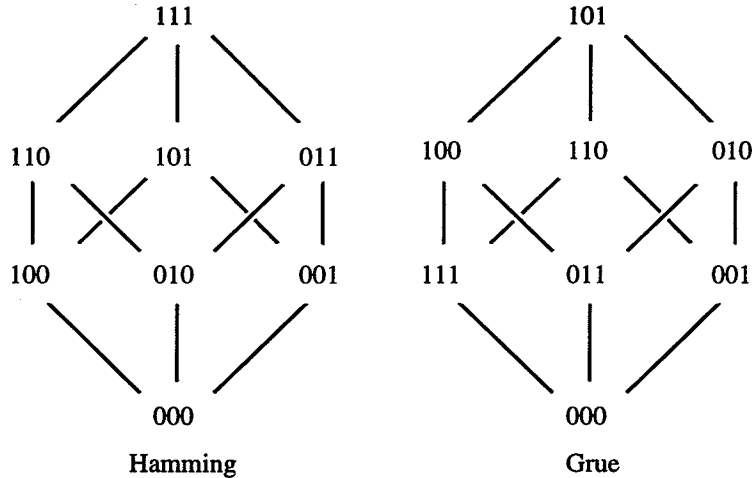
15

Figure 1: Hamming and Grue algebras

These orderings are isomorphic copies of the inclusion ordering on the power set of ω and hence are isomorphic Boolean algebras, but they label this structure very differently (e.g., adjacent elements of the Goodman algebra are complements in the Hamming algebra). Moreover, $G^\omega(e_0)$ is the union of the finite levels of $r_{e_0}^G$, whereas the union of the finite levels of $r_{e_0}^H$, is just $G_{\text{even}}^\omega(e_0)$:

Proposition 15   *For each* $n, G^n(e_0) = \bigcup_{i \leq n} r_{e_0}^G$.

Proof: direct consequence of proposition 14. ⊣

The method $(r_{e_0}^G, *_{S,2})$ identifies $G^\omega(e_0)$ in an intuitively attractive manner. It starts out assuming that the true data stream is e_0. When it encounters a surprise at stage n, it then assumes that the true data stream is $e_0 \updownarrow n$, and so forth, always introducing the smallest number of grue applications to e_0 that is consistent with the data (propositions 21.2 and 20). Recall that $*_{S,2}$ has the objectionable property that a possibility can become more plausible when it is refuted if only very implausible worlds are refuted by the current datum. The implausibility ranking based on grue distance prevenst this possibility from ever occurring. This initial state has the property that, at each stage prior to convergence, a highly plausible (degree 0 or 1) possibility is refuted. Since $\alpha = 2$, all refuted possibilities are pushed up at least one step by $*_{S,2}$. When $\alpha < 2$, refuted possibilities do not rise when the agent's beliefs are not refuted, so the same argument does not work and in fact cannot be made to work since $G^1(e_0)$ is not identifiable by $*_{S,1}$.

Turning to the negative results, it is remarkable that $*_M$, $*_{S,1}$ and $*_{A,1}$ cannot even identify $G^1(e_0)$ (proposition 29). $*_{DP,1}$ survives just one level higher, failing on $G^2(e_0)$ (proposition 28). $*_{G,\alpha}$ compares unfavorably with $*_{S,\alpha}$ and $*_{DP,\alpha}$, because $*_{G,n+1}$ fails on $G^{n+1}(e_0)$, for each n, whereas $*_{S,2}$ and $*_{DP,2}$ succeed on $G^\omega(e_0)$. By proposition 11, $G^n(e_0)$ can be solved with just n retractions by the obvious method that starts out conjecturing e_0 and that refuses to believe in grue operations until they are observed. The negative results imply that this sensible behavior cannot be obtained from $*_M$, $*_{S,1}$ or $*_{A,1}$, no matter how cleverly the initial epistemic state is arranged.

By the following proposition, all of the negative results in the table are instances of inductive amnesia.

---

[21] This idea is familiar in computer science as a way to compress image files. Instead of recording the intensity of each pixel separately, one records the places at which intensity changes, which saves space if many adjacent pixels have the same intensity.

16

Proposition 16 *Let $e_0 \in B$.*

   1. *All of the methods under consideration can remember the past.*

   2. *All of the methods under consideration can project $G^\omega(e_0)$ so long as $\alpha > 0$ Among these, only M fails to be a complete projector.*

Proof: propositions 19, 21, and 10. ⊣

The inductive amnesia results reflect a fundamental, epistemic dilemma for iterated belief revision methods. Recall tha belief revision theory can be stretched in two directions. Lumping all possible worlds together at one level of implausibility makes belief revision behave like an accretive tabula rasa that takes no inductive risks and never encounters a contradiction. Spreading worlds out at distinct levels of implausibility makes belief revision look more like Popper's methodology of bold conjectures and refutations. The former extreme secures perfect memory and the latter secures predictive reliability. But if as α is set too low, the Popperian belief revisor "leaks" information about past refutations since refuted possibilities are inserted back into the enumeration of "live" possibilities. So memory requires a sufficiently "compressed" epistemic state and prediction demands a sufficiently "rarified" epistemic state. Once this fundamental duality is recognized, the question is whether there exists a *critical value* of α below which the competing pressures can no longer be satisfied jointly. Perhaps the most striking result of this investigation is that the methods $*_{S,\alpha}$ and $*_{DP,\alpha}$ enjoy an *infinite* jump in reliability when α is incremented from *one* to *two*. For $\alpha \geq 2$, the methods succeed over the entire grue hierarchy. For $\alpha \leq 2$, neither can cope with more than two grue shifts in the data stream.

0.8 Dropping Well-ordering

So far, it has been assumed that epistemic states well-order the possible worlds in their domains, since epistemic states assume ordinal values. This assumption is not generally accepted in the belief revision community, so we should consider what happens when it is dropped. The positive identifiability results are not affected since success based on a well-ordered initial state is still success. It is the negative results that require scrutiny. For example, well-ordering is involved in the proof that $*_M$ fails on the easy problems $G^1_{even}(e_0)$ and $G^1(e_0)$. Since $*_M$ has a straightforward extension to a wide class of non-well-ordered epistemic states (Boutilier 94), we should examine whether its modest learning abilities improve in this more general setting.

Let $R = (D, \leq)$ be a totally ordered set. Let $\min(R, E)$ denote the set of all minimial elements of $E \cap D$. For present purposes, an *epistemic state* is a total order $R = (D, \leq)$ such that $D \subseteq U$ and for each proposition $E \in \{U\} \cup \{[i,k] : i, k \in \omega\}$, $\min(R, E) \neq \emptyset$. In other words, an epistemic state is a total order on data streams that has a least element and in which each observation of an outcome (consistent with the domain of the order) has a least element. The associated belief state of R is given by $b(R) = \min(R, U)$. Upon receiving new information $[i,k]$, $*_M$ updates the epistemic state $R_1 = (D_1, <_1)$ producing a new state $R_2 = (D_2, <_2)$, such that

   1. $\min(R_2, U) = \min(R_1, [i,k])$ (which is nonempty) and

   2. for all $e, e' \in D_1 - \min(R_1, [i,k])$, $e \leq_2 e' \leftrightarrow e \leq_1 e'$.

In other words, $*_M$ brings $\min(R_1, [i,k])$ to the bottom of the new state and rigidly raises all the rest of the worlds so that the lowest are immediately above $\min(R_1, [i,k])$ in the revised ordering. A inductive method implementing $*_M$ is a pair of form $(R, *_M)$. Now we have:

Proposition 17 *Redefine $*_M$ and epistemic states as just described. Then:*

   1. *$G^1(e_0), G^1_{even}(e_0)$ are identifiable by $*_M$.*

17

2. $G^2(e_0), G^2_{even}(e_0)$ *are not identifiable by* $*_M$.

Proof: in Appendix IV. ⊣

So the learning power of $*_M$ improves slightly in the more general setting in which the well-ordering assumption is dropped. This result illustrates how learning theoretic analysis can be employed to criticize controversial assumptions about the nature of epistemic states.

The well-ordering assumption is also involved in the negative results concerning $*_{A,n}$,[22] $*_{S,1}$, and $*_{DP,1}$. But these methods were originally defined only on ordinal-valued epistemic states (Goldszmidt and Pearl 94; Darwiche and Pearl 97; Spohn 88), and it is unclear how these methods should be extended to arbitrary, totally ordered states. If, more modestly, degrees of implausibility are taken to be numbers in a non-well-ordered system (e.g., rationals, reals, or nonstandard reals) then all of these methods are enumerate-and-test operators if the domain of the initial epistemic state is confined to the $[0,1)$ interval, and hence are empirically complete (proposition 5). A systematic survey of possible generalizations of these methods lies beyond the scope of this paper.

0.9   Conclusion

The normative principles of belief revision theory have been motivated by intuition, coherence, and symmetry considerations. The natural question whether following such a rule would help or hinder the formulation of informative, true beliefs has largely been ignored. Once this question is entertained, a range of interesting and unanticipated issues emerges, such as (i) inductive amnesia, (ii) the essential tension between compression and rarefaction in the epistemic state, (iii) the pivotal significance for the resolvability of this tension of the value $\alpha = 2$, (iv) the idea of generating epistemic states from operations on data streams or as ranks in Boolean algebras, (v) the utility of grue distance for improving the reliability of belief revision methods, (vi) the appealing portrayal of induction as rigid rotation of a hypercube, (vii) the image of tail reversals in data streams "derailing" this rotation and (viii) the relevance for reliability of well-ordered degrees of implausibility. These issues are not drawn from a priori intuitions. They are rigorously derivable from the straightforward aim of arriving at nontrivial, true beliefs. As such, they can serve as well- motivated constraints on theories of rational belief revision.

The results of this study should be expanded and generalized. It is left open, for example, whether $*_{S,2}$ and $*_{DP,2}$ are complete architectures for partition identification. The rest of the results could be cast in more general settings, in which the order of the data may be scrambled, experimental acts may be performed, meaning shifts are possible, and so forth.[23] But even the simplified, narrow setting of the present study illustrates how a systematic logical analysis founded on the aim of finding true, informative beliefs can serve as a powerful and interesting constraint on belief revision theorizing; a constraint that, it is hoped, will become as familiar to belief revision theorists as the usual representation and equivalence results are today.

0.10   Acknowledgements

---

[22] Ordinal-valued implausibilities are required for the proofs of propositions 25 and 27.

[23] (Kelly 1996) provides illustrations of how such generalizations might proceed. Osherson and Martin (1995, 1996) develop a reliability analysis of belief revision operators using logical languages.

18

0.11 Bibliography

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

Horacio Arlo Costa (1997). "Belief Revision Conditionals: Basic Iterated Systems", *Annals of Pure and Applied Logic*, forthcoming.

Adnan Darwiche and Judea Pearl (1997) "On the Logic of Iterated Belief Revision", 89:1-29.

John Earman (1992). *Bayes or Bust*, Cambridge: M.I.T. Press.

Peter Gärdenfors (1988) *Knowledge in Flux*, Cambridge: MIT Press.

Stephen Murray Glaister (1997), "Symmetry and Belief Revision", Unpublished Manuscript.

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

Moises Goldszmidt and Judea Pearl (1994) "Qualitative Probabilities for Default Reasoning, Belief Revision, and Causal Modeling", Technical Report r-161-L, Revision II.

Nelson Goodman (1983) *Fact, Fiction and Forecast*, 4th ed. Cambridge: Harvard University Press.

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

William James (1948) "The Will to Believe", in *Essays in Pragmatism*, ed. A. Castell. New York: Macmillan.

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

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

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

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.

J. Kemeny (1953) "The Use of Simplicity in Induction", *Philosophical Review* 62: 391-408.

Thomas Kuhn (1970), *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.

Isaac Levi (1996), *The Enterprise of Knowledge*, Cambridge: M.I.T. Press.

Isaac Levi (1996), *For the Sake of the Argument: Ramsey Test Conditionals, Inductive Inference, and Nonmonotonic Reasoning*, Cambridge: Cambridge University Press.

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.

Eric Martin and Daniel Osherson (1996). "Scientific Discovery Based on Belief Revision", *Journal of Symbolic Logic*, forthcoming.

19

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

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

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

Hilary Putnam (1963) "'Degree of Confirmation' and Inductive Logic", in *The Philosophy of .Rudolph Carnap*, ed. A. Schilpp. LaSalle, Ill.: Open Court.

Frank P. Ramsey (1990) *Philosophical Papers*, H. A. Mellor, ed., Cambridge: Cambridge University Press.

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

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.

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

0.12   Appendix I: Proofs of Characterization Theorems

Proof of proposition 2: $1 \Rightarrow 3$ follows from proposition 1.

$3 \Rightarrow 1$: Suppose K is countable. Then $\Theta_0[K]$ has countably many cells. Moreover, each cell $\{e\} \in \Theta_0[K]$ is K-closed, since $U - \{e\} = \bigcup\{[\epsilon] : \epsilon \not\subseteq e\}$. Hence, by proposition 8, Θ_0 is identifiable given K, so K is identifiable.

$1 \Rightarrow 2$ is immediate.

$2 \Rightarrow 1$: suppose that method f projects K. Now define method g, which identifies K, as follows.

$$g(\epsilon) = \begin{cases} [\epsilon] \cap [\mathrm{lh}(\epsilon)|e] & \text{if } \exists e, \emptyset \neq f(\epsilon) \subseteq [\mathrm{lh}(\epsilon)|e] \\ \{e'\} & \text{otherwise,} \end{cases}$$

where e' is an arbitrary element of $[\epsilon]$.⊣

Proof of proposition 3: ($\rightarrow$): Suppose

(i) f identifies Θ given K with n retractions. Define for each $e \in K$:

(ii) $r(e) = k \Leftrightarrow \mathrm{retractions}(f, e) = k$. Thus,

(iii) $\mathrm{rng}(r) = \{0, \ldots, n\}$ and $K = \mathrm{dom}(r)$.

Let $C \in \Theta$. (1) Suppose for reductio that C is not open in $r^{-1}(k)$. Then some $e \in K \cap r^{-1}(k) \cap C$ is a limit point of $K \cap r^{-1}(k) - C$. So for each i there is an $e_i \in r^{-1}(k) - C$ such that $e_i|i = e|i$. Let w = the least m, such that for all $m' \geq m, f(e|m') = f(e|m)$, so by (i), $f(e|w) \subseteq C$ and (by ii) f uses a full k retractions along e by stage w. But since $e_w \notin C$ and $e_w|w = e|w$, there is a $w' > w$ such that $f(e_w|w) \subseteq C$ but $f(e_w|w') \subseteq K - C$. Hence, f performs more than k retractions along e_w. Contradiction. So

(iv) C is open in $r^{-1}(k)$.

(2) Define:

$G = \{\epsilon \in \omega^{<\omega} : f \text{ performs at least } k+1 \text{ retractions by the end of } \epsilon\}$. Then

$\bigcup_{i=1}^{k} r^{-1}(i) = \{e \in \mathrm{dom}(r) : \mathrm{retractions}(f, e) \leq k\} = K - \bigcup\{[\epsilon] : \epsilon \in G\}$. Hence,

(v) $\bigcup_{i=1}^{k} r^{-1}(i)$ is closed in $\mathrm{dom}(r)$. The proposition is established by (iii), (iv) and (v).

($\leftarrow$): Deferred to proposition 9. $\dashv$

Proof of proposition 4: ($\rightarrow$): let r be as guaranteed by proposition 3. So $K \subseteq \text{dom}(r)$, $\text{rng}(r) = \{0, \ldots n\}$, and for each $e \in K$, $\{e\}$ is open in $r^{-1}(i)$ and for each $k \leq n$, $\bigcup_{i=1}^{k} r^{-1}(i)$ is closed in $\text{dom}(r)$. Let $e \in K$. So there exists a k such that $r(e) = k$. Since $\{e\}$ is open in $r^{-1}(k)$, $\exists n \forall m \geq n, [e|m] \cap r^{-1}(k) = \{e\}$. Also, $\bigcup_{i=1}^{k-1} r^{-1}(i)$ is closed in $\text{dom}(r)$. So $\exists n' \forall m \geq n', [e|m] \cap \bigcup_{i=1}^{k-1} r^{-1}(i) = \emptyset$. So $\forall m \geq \max(n, n'), [e|m] \cap \bigcup_{i=1}^{k} r^{-1}(i) = \{e\}$. Hence e is isolated in $[e]_r^{\leq}$.

($\leftarrow$): Deferred to proposition 9. $\dashv$

Proof of proposition 11: (1) For the $G^n(e_0)$ case, let $r^{-1}(i) = G^i(e_0)$, for each $i \leq n$. For the $G^n_{\text{even}}(e_0)$ case, let $r^{-1}(i) = G^i(e_0)\text{even}(e_0)$. Observe that for each $e \in r^{-1}(i)$, $\{e\}$ is isolated in $[e]_r^{\leq}$ and apply proposition 4.

For the negative claim, assume for reductio that g succeeds on $G^{n+1}(e_0)$ with n mind-changes. Feed to g data drawn from e_0 until a stage k_0 is reached at which g outputs $\{e_0\}$, which g must do since $e_0 \in G^n(e_0)$. Then proceed by feeding $e_1 = e_0 \ddagger k$ until g outputs $\{e_1\}$. This procedure can be continued until $n + 1$ grue operations have been applied. But then $n + 1$ retractions are perfomed by g on the resulting data stream e_{n+1}. Contradiction. The negative argument for $G^n_{\text{even}}(e_0)$ is similar except that we let $e_{i+1} = (e_i \ddagger k) \ddagger k + 1$.

(2) The following method identifies $G^\omega(e_0)$: enumerate the whole set and output the first data stream in the set that is consistent with the finite outcome sequence ϵ seen so far. The negative claim follows from part (1). $\dashv$

0.13 Appendix II: Completeness Proofs

Proposition 18 *For each e, $\{e\}$ is identifiable with 0 retractions by $*_M, *_{A,0}, *_{S,0}, *_{DP,0}$.*

Proof: Choose r so that $\text{dom}(r) = \{e\}$ and $r(e) = 0$. $\dashv$

Proposition 19 *Let $e_0 \in B$. $*_M$ can project $G^\omega(e_0)$.*

Proof: Let $r^{-1}(0) = \{e_0\}$ and let $r^{-1}(1) = \{\neg e_0\}$. If $e \in G^\omega(e_0)$, then e is a finite variant of either e_0 or $\neg e_0$. It is easy to check that $(r, *_M)$ succeeds. $\dashv$

Proof of proposition 5: Suppose Θ is identifiable given K. By proposition 1, we may suppose that for each cell $C_j \in \Theta[K]$, there exists a countable union $B_j = \bigcup_i^\infty S_i^j$ of K-closed sets such that $B_j \cap K = C_j \cap K$. Enumerate $\{S_i^j \cap K : i, j \in \omega\}$ as $R_0, \ldots, R_n, \ldots$. Define
(i) $r(e) = (\mu i)(e \in R_i')$.
Evidently, $\text{rng}(r) \subseteq \omega$. Let $*$ be such that $(r, *)$ enumerates and tests. Hence, $(r, *)$ is locally consistent, positively invariant, and satisfies positive precedence. It remains to show that $(r, *)$ identifies Θ given K. Let $e \in K$. Since the R_i cover K, $r(e)$ is defined and $e \in R_{r(e)}$. Since each R_i is K-closed, so is $[e]_r^{\leq} = \bigcup_{i=0}^{r(e)-1} R_i$, since finite unions of K-closed sets are K-closed. So $U - [e]_r^{\leq}$ is K-open. So there is a set S of finite sequences such that $K - [e]_r^{\leq} = K \cap \bigcup_{\epsilon \in S} [\epsilon]$. Since $e \in K - [e]_r^{\leq}$, there is a k such that $e|k \in S$. So since $\text{dom}(r) = K$, we have for each $k' \geq k$,
(ii) $[e]_r^{\leq} \cap [e|k'] = [e]_r^{\leq} \cap [e|k'] = \emptyset$.
By (ii) and positive precedence, we have that $e \in \text{dom}(r * [[e|k']])$ and for each $e' \in [e]_r^{\leq}$,
(iii) either $e' \notin \text{dom}(r * [[e|k']])$ or $(r * [[e|k']])(e) < r * [[e|k']])(e')$
Since $e \in [e|k']$, we have by positive order-invariance that for each $e'' \in \text{dom}(r * [[e|k']]) - [e]_r^{\leq}$,
(iv) $(r * [[e|k']])(e) < (r * [[e|k']])(e'')$. But by local consistency,
(v) $(r * [[e|k']])^{-1}(0) \neq \emptyset$.
By (iii-v), we have
(vi) $e \in b(r * [[e|k']]) = (r, *)(e|k')$. It remains only to establish that
(vii) $b(r *_{CL} [[e|k']]) \subseteq C$.

21

Let C be the cell of Θ to which e belongs. Let $e' \in \operatorname{dom}(r) - C$. Then $e' \notin R_{r(e)} \subseteq C$, so

(viii) $r(e) \neq r(e')$.

Case I: suppose $e' \notin [e|k']$. Then by positive precedence, either $e' \notin \operatorname{dom}(r * [[e|k']])$ or $(r * [[e|k']])(e) < (r * [[e|k']])(e')$. Thus, $e' \notin b(r *_{CL} [[e|k']]) = (r, *)(e|k')$.

Case II: suppose $e' \in [e|k']$. Then by (ii), and (vii), $r(e) > r(e')$. By positive invariance, $(r *_{CL} [[e|k']])(e') > (r *_{CL} [[e|k']])(e)$, so again $e' \notin b(r *_{CL} [[e|k']]) = (r, *)(e|k')$. $\dashv$

Proof of proposition 8: The implicitations $2 \Rightarrow 1, \ldots, 6 \Rightarrow 1$ are immediate. The implications $1 \Rightarrow 2, 3, 4, 5, 6$ follow from propositions 5 and 7. $\dashv$

Proof of proposition 9: Suppose Θ is identifiable given K with at most n retractions. Then by proposition 3, there exists an r such that

(i) $\operatorname{rng}(r) \subseteq \{0, \ldots n\}$ and

(ii) $K \subseteq \operatorname{dom}(r)$ and for each cell $C \in \Theta$, for each $k \leq n$,

(iii) C is open (and hence clopen) in $r^{-1}(k)$ and

(iv) $\bigcup_{i=1}^{k} r^{-1}(i)$ is K-closed. Let $*$ range over $*_C$, $*_L$, $*_{DP,k+1}$, $*_{S,k+1}$, $*_{A,k+1}$. By (i), $k + 1 \geq r_{\text{above}}(\operatorname{dom}(r))$, so by proposition 7, $*$ generates and tests. So by proposition 5, $(r, *)$ identifies Θ given K.

Let $e \in K$. It remains to show that each of these operators performs at most k retractions along e when started out on r. Suppose that $(r, *)(e|k) \nsubseteq (r, *)(e|k+1)$. Then $b(r * [[e|k+1]]) \nsubseteq b(r * [[e|k]])$. So there exists an e' such that

(v) $(r * [[e|k+1]])(e') = 0$ but $(r * [[e|k]])(e') \neq 0$. So by the definition of $*$, $(r * [[e|k]])(e'|[k, e(k)]) = 0$. So

(vi) $-\min\{(r * [[e|k]])(e'') : e'' \in \operatorname{dom}(r * [[e|k]]) \cap [k, e(k)]\} + (r * [[e|k]])(e') = 0$. By (v) and (vi), $\min\{(r * [[e|k]])(e'') : e'' \in \operatorname{dom}(r * [[e|k]]) \cap [k, e(k)]\} > 0$. Hence, $-\min\{(r * [[e|k]])(e'') : e'' \in \operatorname{dom}(r * [[e|k]]) \cap [k, e(k)]\} + (r * [[e|k]])(e) < (r * [[e|k]])(e)$, so $(r * [[e|k+1]])(e) < (r * [[e|k]])(e)$. So we have that

(vii) for each k such that $(r, *)(e|k) \nsubseteq (r, *)(e|k+1)$, $(r * [[e|k+1]])(e) < (r * [[e|k]])(e)$. But by hypothesis, $r(e) \leq n$. Hence, $(r, *)$ performs at most n retractions along e. $\dashv$

Proof of proposition 10: $3 \Rightarrow 1, \ldots, 7 \Rightarrow 1, 8 \Rightarrow 2, 9 \rightarrow 2$ are immediate.

$1 \Rightarrow 3, \ldots, 1 \Rightarrow 6$ are instances of proposition 8. $1 \Leftrightarrow 2$ is from proposition 2.

$2 \Rightarrow 9$: Let $* = *_{S,1}$. Let K be projectable. Then K is countable. Enumerate K as $e_0, e_1, \ldots$. Let $r^{-1}(i) = \{e_i\}$. Let $e \in K$, so for some $i, e = e_i$. First it is established that:

(i) $\forall n \forall i, (r *_{S,1} [[e|n]])^{-1}(i)$ is finite.

This is evident by the definition of r when $n = 0$. Suppose (i) holds up to n. Let $m = \min\{(r * [[e|n]])(e') : e' \in \operatorname{dom}(r * [[e|n]]) \wedge e' \in [n, e(n)]\}$ and let

$m' = \min\{(r * [[e|n]])(e') : e' \in \operatorname{dom}(r * [[e|n]]) \wedge e' \in U - [n, e(n)]\}$

Then by the definition of $*$, we have $(r * [[e|n+1]])^{-1}(i) = ([n, e(n)] \cap (r * [[e|n]])^{-1}(m+i)) \cup ((U - [n, e(n)]) \cap (r * [[e|n]])^{-1}(m'+i-1))$, under the convention that $(r * [[e|n]])^{-1}(z)) = \emptyset$ if $z < 0$. This set is finite by the induction hypothesis. So we have (i). Next, we establish

(ii) If $b(r * [[e|n]]) \cap [n, e(n)] = \emptyset$ then $(r * [[e|n+1]])(e) \leq (r * [[e|n]])(e) - 1$.

For suppose $b(r * [[e|n]]) \cap [n, e(n)] = \emptyset$. Then since $e \in [n, e(n)]$, we have $(r * [[e|n+1]])(e) =$

$= -\min\{(r * [[e|n]])(e') : e' \in \operatorname{dom}(r * [[e|n]]) \cap [n, e(n)]\} + (r * [[e|n]])(e)$

$\leq -1 + (r * [[e|n]])(e)$. So we have (ii). Next we establish:

(iii) if $\forall e' \in b(r * [[e|n]]), n|e' \neq n|e$, then $\exists m \geq n, b(r * [[e|m]]) \cap [m, e(m)] = \emptyset$.

For suppose that for all $e' \in b(r * [[e|n]]), n|e' \neq n|e$. Suppose for reductio that for all $m \geq n$, $b(r * [[e|m]]) \cap [m, e(m)] \neq \emptyset$. Then by timidity and stubbornness (proposition 7),

(iv) $\forall m \geq n, b(r * [[e|m]]) = b(r * [[e|n]]) \cap [n, e(n)] \cap \ldots \cap [m-1, e(m-1)]$. $b(r * [[e|n]])$ is finite by (i). So by the hypothesis of (iii), there exists an $m' \geq n$ such that $b(r * [[e|n]]) \cap [n, e(n)] \cap \ldots \cap [m-1, e(m-1)] = \emptyset$. By (iv), $b(r * [[e|m']]) = \emptyset$, contradicting local consistency and establishing (iii). Next we need

(v) if $\exists e' \in b(r * [[e|n]])$ such that $n|e' = n|e$, then then $\exists m \geq n$ such that $b(r * [[e|m]]) = \{e'' \in b(r * [[e|n]]) :$
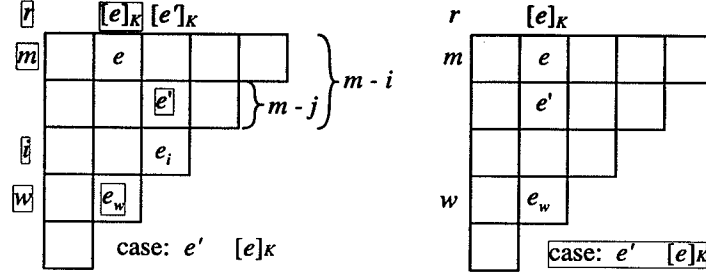
Figure 2: Completeness of $*_{DP,2}$

$m|e'' = m|e\}$. For by (i), there is an $m \geq n$ such that $b(r * [[e|n]]) \cap [n, e(n)] \cap \ldots \cap [m-1, e(m-1)] = \{e'' \in b(r * [[e|m]]) : m|e = m|e''\}$. But by timidity and stubbornness, $b(r * [[e|m]]) = b(r * [[e|n]]) \cap [n, e(n)] \cap \ldots \cap [m-1, e(m-1)]$, establishing (v). Finally, it is shown that

(vi) if $\forall e' \in b(r * [[e|n]]), n|e' = n|e$, then $\forall m \geq n, b(r * [[e|m]]) = b(r * [[e|n]])$.

For by local consistency, $b(r * [[e|n]]) \neq \emptyset$. So by timidity and stubbornness, (iv) holds at each stage $m \geq n$, yielding (vi).

Consider the following procedure: Start out at stage 0 with r and let $n_0 = 0$.

At stage k, if $b(r)$ contains no e' such that $n_k|e' = n_k|e$, apply (iii) to obtain an n_{k+1} such that $b(r * [[e|n_{k+1}]]) \cap [n_{k+1}, e(n_{k+1})] = \emptyset$. Otherwise stop the procedure.

The procedure halts by stage $r(e)$, for by (ii), $(r * [[e|n_{k+1}]])(e) \leq (r * [[e|n_k]])(e) - 1$ (i.e., e drops by at least one step at each stage) and when $e \in b(r * [[e|n_{r(e)}]])$, the condition for continuing is no longer satisfied. Let k be the last stage and let $m = n_k$. Then by the halting condition, we have $b(r)$ contains an e' such that $m|e' = m|e$. By (v), there is an $m' \geq m$ such that $\emptyset \subset b(r * [[e|m']]) \subseteq [m'|e]$ By (vi), this situation remains for each $m'' \geq m'$. So $(r, *)$ projects K.

$2 \Rightarrow 8$: Follow the steps in the preceding argument. A shorter argument may be given using the climbing lemma.

$1 \Rightarrow 7$: Define restricted Hamming distance as follows:

$\rho_k(e, e') = |\Delta(e, e') \cap \{0, \ldots, k-1\}|$. While this is not a metric, it does satisfy the triangle inequality, which will be used below.

Suppose K is identifiable. So by proposition 2, K is countable. If $e \in K$ then let $[e]_K$ be the set of all finite variants of e in K. Since K is countable, we may enumerate these classes as $C_0, \ldots, C_n, \ldots$. For each i, choose a unique element $e_i \in C_i$. For each $e \in K$, let $z(e)$ denote the unique w such that $e \in C_w$. Now define the IA r as follows:

$r(e) = \rho(e_{z(e)}, e) + z(e)$. Let $e \in K$ and let

(i) $r(e) = m$ and $z(e) = w$. Define $P = \{i \leq m : i \neq w\}$. If $i \in P$, then there are infinitely many m such that $e_i(m) \neq e(m)$, so there is a k_i such that $\rho_{k_i}(e_i, e) > 2m$. Moreover, there is a j sufficiently large so that $\rho_j(e_w, e) = \rho(e_w, e)$. Since P is finite, let $k = \max(\{k_i : i \in P\} \cup \{j\})$. Let $k' \geq k$. So

(ii) $\rho_{k'}(e_i, e) > 2m$. We now establish that

(iii) $\forall k' \geq k, e' \in K, e' \neq e \Rightarrow (r *_{DP,2} [[e|k']])(e') > (r *_{DP,2} [[e|k']])(e)$. Let $e' \in K$, $e' \neq e$.

Case 1: $e' \in [e]_r^{\leq} - [e]_K$. So $z(e') \neq w$. Let $z(e') = i$. So by the definition of r,

(iv) $\rho_{k'}(e_i, e') \leq \rho(e_i, e') \leq m$. By the triangle inequality:

$\rho_{k'}(e, e') + \rho_{k'}(e_i, e') \geq \rho_{k'}(e_i, e)$, so

$\rho_{k'}(e, e') \geq \rho_{k'}(e_i, e) - \rho_{k'}(e_i, e') > 2m - m = m$, by (ii, iv). Hence

(v) $\rho_{k'}(e, e') > m$. By the climbing lemma (proposition 6),

$(r *_{DP,2} [[e|k']])(e) - (r *_{DP,2} [[e|k']])(e') \leq (r(e) - r(e')) - 2\rho_{k'}(e, e')$

$< m - r(e') - 2m \leq 0$ (by i, v), so (iii) obtains in this case.

Case 2: $e' \in [e]^{\leq}_{\hat{r}} \cap [e]_K$. By choice of k,

(vi) $\rho_{k'}(e_w, e) = \rho(e_w, e)$. By the triangle inequality,

$\rho_{k'}(e, e') + \rho_{k'}(e_w, e) \geq \rho_{k'}(e_w, e)$, so

(vii) $\rho_{k'}(e, e') \geq \rho_{k'}(e_w, e) - \rho_{k'}(e_w, e')$. By the definition of r:

$r(e) - r(e') = (\rho(e_w, e) + w) - (\rho(e_w, e') + w)$

$= \rho(e_w, e) - \rho(e_w, e')$

$\leq \rho_{k'}(e_w, e) - \rho_{k'}(e_w, e')$ (by vi)

$\leq \rho_{k'}(e, e')$ (by vii). So

(viii) $r(e) - r(e') \leq \rho_{k'}(e, e')$. By proposition 6,

$(r *_{DP,2} [[e|k']])(e') - (r *_{DP,2} [[e|k']])(e) \leq (r(e') - r(e)) - 2\rho_{k'}(e, e')$

$\leq \rho_{k'}(e, e') - 2\rho_{k'}(e, e')$ (by viii), which quantity is negative, so long as $\rho_{k'}(e, e') > 0$.[24] So it suffices for (iii) to show that $\rho_{k'}(e, e') > 0$. Suppose $\rho_{k'}(e, e') = 0$. Then

(ix) $e|k' = e'|k'$. By (vi), we have

(x) $k'|e = k'|e_w$. By the case hypothesis, $r(e) \geq r(e')$. So by the definition of r,

$\rho(e_w, e) + w \geq \rho(e_w, e') + w$, so

$\rho(e_w, e) \geq \rho(e_w, e')$. So by (ix, x),

(xi) $k'|e' = k'|e_w$. But by (ix, x, xi), we have $e = e'$, contradicting the choice of e'. Hence, $\rho_{k'}(e, e') > 0$ and we have (iii) under this case.

Case 3: $e' \notin [e]^{\leq}_{\hat{r}}$. Then (iii) follows by positive order-invariance (proposition 7) and proposition 6).

By proposition 7, $r*_{DP,2}$ is locally consistent. Hence,

(xii) for each $k' \geq k, b(r *_{DP,2} [[e|k']]) \neq \emptyset$. So by (iii), we have that for each $k' \geq k, b(r *_{DP,1} [[e|k']]) = \{e\}$. ⊣

0.14 Appendix III: A Positive Result for S,2

This appendix is devoted to proving that $*_{S,2}$ identifies $G^\omega(e_0)$.

Definition 17 *Let α be an ordinal and let $e_0 \in B$. Let $r = r^G_{e_0}$. Let ϵ be a finite boolean sequence of nonzero length and let $last(\epsilon)$ denote the last item occurring in ϵ. Then define:*

1. $\beta_\alpha(\epsilon, b) = \begin{cases} -1 & \text{if } b = last(\epsilon) \wedge lh(\epsilon) - 1 \in \Gamma(e_0, \epsilon) \\ 0 & \text{if } b = last(\epsilon) \wedge lh(\epsilon) - 1 \notin \Gamma(e_0, \epsilon) \\ \alpha & \text{if } b \neq last(\epsilon) \wedge lh(\epsilon) - 1 \in \Gamma(e_0, \epsilon) \\ \alpha - 1 & \text{if } b \neq last(\epsilon) \wedge lh(\epsilon) - 1 \notin \Gamma(e_0, \epsilon) \end{cases}$

2. $\beta'_\alpha(\epsilon, e') = \Sigma^{lh(\epsilon)}_{i=1} \beta(\epsilon|i, e'(i-1))$.

3. $\beta^r_\alpha(\epsilon, e') = r(e') + \beta'_\alpha(\epsilon, e')$.

Proposition 20 *Let e_0, r be as in the preceding definition. Let $e, e' \in G^\omega(e_0)$. Let $e[m] = e_0 \ddagger \{i < m : i \in \Gamma(e_0, e)\}$ and let $\alpha \geq 2$. Then*

1. $\beta^r_\alpha(e|m, e') \geq 0$.

2. If $e|m = e'|m$ then $\beta^r_\alpha(e|m, e') = |\{i \geq m : i \in \Gamma(e_0, e')\}|$.

3. If $e' \neq e[m]$ then $\beta^r_\alpha(e|m, e') > 0$.

---

[24]For those tracing the magic of $\alpha = 2$, note that the argument would fail here if $\alpha = 1$.

Proof: Define
$M = \{i \in \omega : i < m\}$,
$G = \Gamma(e_0, e)$,
$G' = \Gamma(e_0, e')$, and
$E = \{i \in \omega : e(i) = e'(i)\}$.
(1) using the definition of β_α and the fact that $\alpha \geq 2$, we have:
$\beta_\alpha^r(e|m, e') = r(e') + \beta_\alpha'(e|m, e') =$
$= r(e') + \Sigma_{i \in M \cap G \cap E} \beta_\alpha(e|i+1, e'(i)) +$
$+ \Sigma_{i \in (M-E) \cap G} \beta_\alpha(e|i+1, e'(i)) +$
$+ \Sigma_{i \in (M \cap E) - G} \beta_\alpha(e|i+1, e'(i)) +$
$+ \Sigma_{i \in (M-E) - G} \beta_\alpha(e|i+1, e'(i)) \geq$
$\geq |G'| - |M \cap G \cap E| + |(M-E) \cap G| + 0 + |(M-E) - G| =$
$= |G'| - |M \cap G \cap E| + |M - E| =$
$= |G'| - |M \cap G \cap E \cap G'| + |M - E| - |M \cap G \cap E - G'| \geq$
$\geq |M - E| + |(M \cap G \cap E) - G'|$, so it suffices to show that $|M - E| \leq |(M \cap G \cap E) - G'|$. For this we
construct an injection f from $|(M \cap G \cap E) - G'|$ to $|M - E|$. Let $i \in (M \cap G \cap E) - G'$. So we have
(i) $i < m$, (ii) $e(i) = e'(i)$, (iii) $i \in \Gamma(e_0, e)$ and (iv) $i \notin \Gamma(e_0, e')$. Suppose for reductio that $i = 0$. Then
by (iv), $e_0(i) = e'(i)$ and by (ii), $e(i) = e'(i)$, so $e_0(i) = e(i)$, contradicting (iii). So we may assume
(v) $i > 0$. Define $f(i) = i - 1$, which is evidently injective and it is also immediate that $f(i) \in M$ if
$i \in M$. Suppose for reductio that $f(i) = i - 1 \in E$, so $e(i-1) = e'(i-1)$. Then by (iii, iv, v), we obtain
$e(i) \neq e_0(i)$, contradicting (iii). Hence, $f(i) \in M - E$.

(2) Suppose $e|m = e'|m$. $r(e') = |G'|$. For each $j \leq m$, if $j \in M \cap G'$, then $\beta_\alpha(e|j+1, e'(j)) = -1$
and if $j \in M - G'$, then $\beta_\alpha(e|j+1, e'(j)) = 0$. Hence, $\beta_\alpha'(e|m, e') = -|M \cap G'|$. So $\beta_\alpha^r(e|m, e') =$
$|G'| - |M \cap G'| = |G' - M| = |\{i \geq m : i \in \Gamma(e_0, e')\}|$.

(3) We begin by establishing
(i) $\beta_\alpha^r(e|m, e') = 0 \Rightarrow M \cap G' \subseteq G$. Suppose for contraposition that there exists $k \in M \cap G' - G$. We will
construct e'' such that $\beta_\alpha^r(e|m, e'') < \beta_\alpha^r(e|m, e')$, so by (1), $\beta_\alpha^r(e|m, e') > 0$.

The construction of e'' proceeds as follows. If $e'(k) = e(k)$, let e'' be just like e' except that $e''(k-1) =$
$\neg e'(k-1)$. Else, e'' is just like e' except that $e''(k) = \neg e'(k)$. This construction is well-defined because
$e'(k) \neq e(k)$ if $k = 0$. For suppose otherwise. Since $k \in G'$, $e'(0) \neq e_0(0)$. But by hypothesis, $e'(0) = e(0)$,
so $e(0) \neq e_0(0)$ and hence $0 = k \in G$, contradicting the choice of k.

We now show that
(ii.a) $r(e'') \leq r(e')$ and
(ii.b) $\beta_\alpha'(e|m, e'') < \beta_\alpha'(e|m, e')$.

Case: $e(k) = e'(k)$. Then $k > 0$ and e'' is just like e' except that $e''(k-1) = \neg e'(k-1)$. So
(iii) $e(k) = e'(k) = e''(k)$. Since $k \in G' - G$, we have
(iv.a) $e(k-1) = e(k) \leftrightarrow e_0(k-1) = e_0(k)$ and
(iv.b) $e'(k-1) \neq e'(k) \leftrightarrow e_0(k-1) = e_0(k)$. So
$e(k-1) \neq e(k) \leftrightarrow e'(k-1) = e'(k)$. So by the case hypothesis,
$e(k-1) \neq e(k) \leftrightarrow e'(k-1) = e(k)$. Hence,
(v) $e''(k-1) = e(k-1) \neq e'(k-1)$. Also, by (iv.b)
(vi) $k \notin G''$. Since e'' differs from e' only at $k-1$, we also have:
(vii) for all $j \notin \{k, k-1\}$, $j \in G'' \leftrightarrow j \in G'$. By (vi, vii), $|G''| \leq |G'|$, which is just (ii.a). By the definition
of β_α, (v), and the fact that $\alpha \geq 2$,
$\beta_\alpha(e|k, e'(k-1)) - \beta_\alpha(e|k, e''(k-1)) \geq \alpha - 1 - 0 \geq 1$.[25] For each $j \neq k-1$, $e'(j) = e''(j)$, so $\beta_\alpha(e|j+1, e'(j)) = \beta_\alpha(e|j+1, e''(j))$, so we have (iv.b).

Case: $e'(k) \neq e(k)$. Then e'' is just like e' except that $e''(k) = \neg e'(k)$. So by the case hypothesis,
(viii) $e''(k) = e(k) \neq e'(k)$. Since $k \in G'$, and $e''(k) = \neg e'(k)$, it follows that

---

[25] $\alpha = 2$ is critical for the argument at this stage.

(ix) $k \notin G''$. Since e'' differs from e' only at k, we also have:

(x) for all $j \notin \{k, k+1\}, j \in G'' \leftrightarrow j \in G'$. By (viii, ix), we have $|G''| \leq |G'|$, which is just (ii.a). By (viii) and the fact that $\alpha \geq 2$,

(xi) $\beta_\alpha(e|k+2, e'(k+1)) - \beta_\alpha(e|k, e''(k-1)) \geq \alpha - 1 \geq 1$.[26] e' and e'' agree everywhere else, so again we have (iv.b).

The next task is to establish:

(xii) $\beta_\alpha^r(e|m, e') = 0 \Rightarrow G' - M = \emptyset$. Suppose that $k \geq m$ and $k \in G'$. So k contributes one unit to $r(e')$. Since $k \geq m$, k contributes nothing to the sum $\beta_\alpha'(e|m, e')$. Let $e'' = e_0 \ddagger (G' - \{k\})$. Then $\beta_\alpha^r(e|m, e'') = r(e'') + \beta_\alpha'(e|m, e'') = r(e') - 1 + \beta_\alpha'(e|m, e') = \beta_\alpha^r(e|m, e') - 1$. So by (1), $\beta_\alpha^r(e|m, e') > 0$.

Finally we show that

(xiii) $\beta_\alpha^r(e|m, e') = 0 \Rightarrow M \cap G \subseteq G'$. Suppose that $\beta_\alpha^r(e|m, e') = 0$. Suppose for reductio that $D = (M \cap G) - G' \neq \emptyset$. By the hypothesis and (i, xiii), we have $G' - M = G - M$ and $G' \cap M \subseteq G \cap M$. So $r(e) - r(e') = |G| - |G'| = |D|$. So if we establish that

(xiv) $\beta_\alpha'(e|m, e') - \beta_\alpha'(e|m, e) > |D|$,

then we have $\beta_\alpha^r(e|m, e') > \beta_\alpha^r(e|m, e)$ so by (1), $\beta_\alpha^r(e|m, e') > 0$. It therefore suffices to establish (xiv).

Let D be enumerated in ascending order as $\{k_1, \ldots, k_d\}$. Observe that $e|k_1 = e'|k_1$ so since $k_1 \in G - G'$, $e(k_1) \neq e'(k_1)$. Thereafter, there is constant disagreement between e and e' until k_2, where another reversal of sense yields constant agreement until k_3, etc. In general, we have for each j such that $1 \leq k \leq d$:

(xv) $e(k_j) = e'(k_j) \Leftrightarrow j$ is even.

Also, we have by the definition of β_α:

(xvi) If $e(k_j) \neq e'(k_j)$ then $\beta_\alpha(e|k_j+1, e'(k_j)) - \beta_\alpha(e|k_j+1, e(k_j)) = \alpha - (-1) \geq 3$ since[27] $\alpha \geq 2$ and

(xvii) If $e(k_j) = e'(k_j)$ then $\beta_\alpha(e|k_j+1, e'(k_j)) - \beta_\alpha(e|k_j+1, e(k_j)) = (-1) - (-1) = 0$.

By (xv, xvi, xvii) we have

(xviii) $\Sigma_{j=1}^d \beta_\alpha(e|k_j+1, e'(k_j)) - \beta_\alpha(e|k_j+1, e(k_j)) \geq 3(d+1)/2$ if d is odd and $\geq 3d/2$ if d is even (note that $3(d+1)/2$ is the number of odd numbers $\leq d$ when d is odd). Observe that

(xix) for all $d > 0$, $3(d-1)/2 > d$ if d is odd and $3d/2 > d$ if d is even.[28] We haven't yet included in the sum terms whose indices are not in D. So let $0 < k \leq m$ and suppose $k - 1 \notin D$. Then by the definition of β_α and the fact that $\alpha \geq 2$, we have

(xx) $\beta_\alpha(e|k, e'(k-1)) - \beta_\alpha(e|k, e(k-1)) \geq 0$. So by (xviii, xix, xx), we have, $\beta_\alpha'(e|m, e') - \beta_\alpha'(e|m, e) > d$, establishing (xiv) and hence (xiii).

Now suppose that $\beta_\alpha^r(e|m, e') = 0$. By (i, xii, xiii) we infer that $e' = e_0 \ddagger \{i > m : i \in \Gamma(e_0, e)\} = e[m]$, which completes the proof of (3). ⊣

Next, we establish that S_α implements β_α^r when $\alpha \geq 2$ and r is the grue ranking.

Proposition 21 *Let $\alpha \geq 2$ and let $e_0, r, e, e', e[n]$ be as in proposition 20 and let $m \in \omega$. Then*

1. $(r *_{S,2} [[e|n]])(e') = \beta_\alpha^r(e|m, e')$,

2. $(r, *_{S,2})(e|n) = \{e[n]\}$, *and*

3. $(r, *_{S,2})$ *identifies* $G^\omega(e_0)$.

Proof: (2) By proposition 20.2, $\beta_\alpha^r(e|n, e[n]) = 0$. By proposition 20.3, for all $e' \neq e[m]$, $\beta_\alpha^r(e|n, e') > 0$. So the result follows from (1).

(3) is a consequence of (2). (2) says that $(r, *_{S,2})$ always remembers the past and conjectures that the current datum will be repeated forever. If $e \in G^\omega(e_0)$ then for some n, $e \in G^n(e_0)$. Let n' be the least such n. Then the method $(r, *_{S,2})$ retracts exactly n' times prior to stabilizing to $\{e\}$.

---

[26] $\alpha = 2$ is critical for the argument at this stage.

[27] $\alpha = 2$ is critical for the argument at this stage.

[28] The inequality is barely strict at $d = 1$ and would fail if $\alpha = 1$, illustrating once again the critical role of the value $\alpha \geq 2$.

(1) By induction on n. Let $* = *_{S,\alpha}$. $\beta_\alpha^r((),e') = r(e') + \beta_\alpha'((),e') = r(e') = (r*())(e')$. Now suppose that for each $e' \in G^\omega(e_0)$, $(r*[[e|n]])(e') = \beta_\alpha^r(e|n,e')$. Then since $\alpha \geq 2$, proposition 20 parts 2 and 3 yield

(i) If $e|n = e'|n$ then $\beta_\alpha^r(e|n,e') = |\{i \geq n : i \in \Gamma(e_0,e')\}|$ and

(ii) If $e' \neq e[n]$ then $\beta_\alpha^r(e|n,e') > 0$. Now consider $(r*[[e|n+1]])(e')$.

Case 1: $e'(n) = e(n)$ Then

$(r*[[e|n+1]])(e') = (r*[[e|n]])(e'|[n,e(n)])$
$= -\min\{(r*[[e|n]])(e'') : e'' \in \mathrm{dom}(r*[[e|n]]) \cap [n,e(n)]\} + (r*[[e|n]])(e')$.

Case 1.A: $n \in \Gamma(e_0,e)$. Hence, $e[n] \notin [n,e(n)]$ (recall that $e[n] = e_0 \ddagger \{i < n : i \in \Gamma(e_0,e)\}$). So by (ii) and the induction hypothesis,

(iii) $0 \notin \{(r*[[e|n]])(e'') : e'' \in \mathrm{dom}(r*[[e|n]]) \cap [n,e(n)]\}$.

Since $e[n+1]|n+1 = e|n+1$ and $\{i \geq n : i \in \Gamma(e_0,e[n+1])\} = \{n\}$, we obtain by (i) and the induction hypothesis that,

(iv) $(r*[[e|n]])(e[n+1]) = 1$. Also,

(v) $e[n+1] \in [n,e(n)]$, since $e[n] \notin [n,e(n)]$ and $n \in \Gamma(e_0,e)$. By (iii, iv, v):

$\min\{(r*[[e|n]])(e'') : e'' \in \mathrm{dom}(r*[[e|n]]) \cap [n,e(n)]\} = 1$. So,

$(r*[[e|n+1]])(e') = (r*[[e|n]])(e') - 1$
$= \beta_\alpha^r(e|n,e') - 1$ (by the induction hypothesis)
$= \beta_\alpha^r(e|n,e') + \beta_\alpha(e|n+1,e'(n))$ (by the case hypotheses)
$= \beta_\alpha^r(e|n+1,e')$.

Case 1.B: $n \notin \Gamma(e_0,e)$. Hence, $e[n] \in [n,e(n)]$. So by (i) and the induction hypothesis,

$\min\{(r*[[e|n]])(e'') : e'' \in \mathrm{dom}(r*[[e|n]]) \cap [n,e(n)]\} = 0$. Hence,

$(r*[[e|n+1]])(e') = (r*[[e|n]])(e') = \beta_\alpha^r(e|n,e') + 0$ (by the induction hypothesis)
$= \beta_\alpha^r(e|n,e') + \beta_\alpha(e|n+1,e'(n))$ (by the case hypotheses)
$= \beta_\alpha^r(e|n+1,e')$.

Case 2: $e'(n) \neq e(n)$. Then

$(r*[[e|n+1]])(e') = (r*[[e|n]])(e'|U - [n,e(n)]) + \alpha$
$= -\min\{(r*[[e|n]])(e'') : e'' \in \mathrm{dom}(r*[[e|n]]) \cap (U - [n,e(n)])\} + (r*[[e|n]])(e') + \alpha$.

Case 2.A: $n \in \Gamma(e_0,e)$. Hence, $e[n] \notin [n,e(n)]$. So by (i) and the induction hypothesis,

$\min\{(r*[[e|n]])(e'') : e'' \in \mathrm{dom}(r*[[e|n]]) \cap (U - [n,e(n)])\} = 0$.

Hence, $(r*[[e|n+1]])(e') = 0 + (r*[[e|n]])(e') + \alpha$
$= \beta_\alpha^r(e|n,e') + \alpha$ (by the induction hypothesis)
$= \beta_\alpha^r(e|n,e') + \beta_\alpha(e|n+1,e'(n))$ (by the case hypotheses)
$= \beta_\alpha^r(e|n+1,e')$.

Case 2.B: $n \notin \Gamma(e_0,e)$. Hence, $e[n] \in [n,e(n)]$. So by (ii) and the induction hypothesis

(vi) $\min\{(r*[[e|n]])(e'') : e'' \in \mathrm{dom}(r*[[e|n]]) \cap U - [n,e(n)]\} > 0$. Let

$e'' = e[n] \ddagger n$. $e''|n = e|n$ and $\{i \leq n : i \in \Gamma(e_0,e'')\} = \{n\}$. So by (i) and the induction hypothesis,

(vii) $(r*[[e|n]])(e'') = 1$. Since $e[n](n) = e(n)$,

(viii) $e'' \notin [n,e(n)]$. Since $e'' \in G^\omega(e_0) = \mathrm{dom}(r)$, (vi, vii, viii) yield

$\min\{(r*[[e|n]])(e'') : e'' \in \mathrm{dom}(r*[[e|n]]) \cap (U - [n,e(n)])\} = 1$. So

$(r*[[e|n+1]])(e') = (r*[[e|n]])(e') + \alpha - 1$
$= \beta_\alpha^r(e|n,e') + \alpha - 1$ by the induction hypothesis
$= \beta_\alpha^r(e|n,e') + \beta_\alpha(e|n+1,e'(n))$ by the case hypotheses
$= \beta_\alpha^r(e|n+1,e')$. $\dashv$

Proof of proposition 13: Let $* = *_{S,2}$. Recall that $(\neg e_0)(n) = \neg(e_0(n))$. Let $r \supseteq r_{e_0}^H|G_{\mathrm{even}}^\omega$ and suppose for reductio that $(r,*_{S,2})$ identifies $G_{\mathrm{even}}^\omega(e_0) \cup \{\neg e_0\}$. For each i, let $e_i = (\neg e_0) \ddagger i = (e_0 \ddagger 0) \ddagger i \in G_{\mathrm{even}}^2(e_0)$.

Case A: Suppose $r(\neg e_0) \geq \omega$. Then for each i, $r(e_i) < r(\neg e_0)$, contradicting the isolation condition (proposition 23).

Case B: So suppose for some $n \in \omega$ that $r(\neg e_0) = n$. By the reductio hypothesis, there is a $k \in \omega$ such that

(i) $(r * [[\neg e_0 | k]])^{-1}(0) = \{\neg e_0\}$. Let $j = \max\{n+1, k\}$. Then

(ii) $r(e_j) > n = r(\neg e_0)$ and $r(e_{j+1}) = r(e_j) + 1$ and

(iii) $e_j | j = e_{j+1} | j = \neg e_0 | j$. By timidity and stubbornness and (i, iii), for each $j' \leq j+1$,

(iv) $(r * [[e_{j+1} | j']])^{-1}(0) = \{\neg e_0\}$. By (iv),

(v) $(r * [[e_{j+1} | j]])(e_j) > 0$. By positive invariance and (ii),

(vi) $(r * [[e_{j+1} | j]])(e_{j+1}) = (r * [[e_{j+1} | j]])(e_j) + 1$. By positive invariance (iv), and (iii),

(vii) $(r * [[e_{j+1} | j+1]])(e_{j+1}) = (r * [[e_{j+1} | j]])(e_{j+1})$. By (iii) and (iv),

(viii) $\min\{(r * [[e_{j+1} | j]])(e') : e' \in \mathrm{dom}(r) \cap (U - [j, e_{j+1}(j)])\} \geq 1$. So since $e_j(j) \neq e_{j+1}(j)$, the definition of $*$ yields

(ix) $(r * [[e_{j+1} | j+1]])(e_j) \leq -1 + (r * [[e_{j+1} | j]])(e_j) + 2$
$\leq (r * [[e_{j+1} | j]])(e_j) + 1$
$\leq (r * [[e_{j+1} | j]])(e_{j+1})$ by (vi). Now $(j+1 | e_j) = (j+1 | e_{j+1})$, so by positive invariance, for all $k' \geq j+1$, $(r * [[e_{j+1} | k']])(e_j) \leq (r * [[e_{j+1} | k']])(e_{j+1})$. Hence, for all such k', $(r * [[e_{j+1} | k']])^{-1}(0) \neq \{e_{j+1}\}$. ⊣

0.15   Appendix IV: Restrictiveness Proofs

Proof of proposition 17: (1) Case $G^1(e_0)$: let $R = (G^1(e_0)), \leq$ be defined so that $b(R) = \{e_0\}$ and for each $k, k' > k \in \omega$, $e_0 \updownarrow k' < e_0$ (note that this condition induces an infinite descending chain in R). It is easy to see that R is an epistemic state and that $(R, *_M)$ succeeds.

Case $G^1_{\mathrm{even}}(e_0)$: Let $R = (G^1_{\mathrm{even}}(e_0)), \leq$ be defined so that $b(R) = \{e_0\}$ and for each $k, k' \geq k, k_1 > k', k_2 > k \in \omega$, $e_0 \updownarrow \{k', k_1\} \leq e_0 \updownarrow \{k, k_2\}$ (this condition also induces an infinite descending chain). R is an epistemic state and $(R, *_M)$ succeeds.

(2) Case $G^1(e_0)$: let $* = *_M$, let $R = (D, <_R)$, and suppose for reductio that $(R, *)$ identifies $G^2(e_0)$. Then for some k,

(i) $b(R * [[e_0 | k]]) = \{e_0\}$. Define $e = e_0 \updownarrow k \mathrm{in} G^2(e_0)$. By proposition 24, we can find $k' > k$ such that, letting $e' = e \updownarrow k' \in G^2(e_0)$,

(ii) $e' >_{(R*[[e_0|k]])} e$. But (i, ii) and the stacking lemma (proposition 22) contradict the reductio hypothesis.

Case $G^1_{\mathrm{even}}(e_0)$: let $* = *_M$ and suppose for reductio that $R = (D, *)$ identifies $G^2_{\mathrm{even}}(e_0)$. Then for some k,

(i) $b(R * [[e_0 | k]]) = \{e_0\}$. Define $e = e_0 \updownarrow \{k, k+1\} \in G^2_{\mathrm{even}}(e_0)$. By proposition 26, we can find $k' > k+1$ such that, letting $e' = e \updownarrow \{k', k'+1\} \in G^2_{\mathrm{even}}(e_0)$,

(ii) $e' >_{(R*[[e_0|k]])} e$. Let $e'' = e_0 \updownarrow \{k', k'+1\} \in G^2_{\mathrm{even}}(e_0)$. By the propping lemma (proposition 22) and (i),

(iii) $e' \geq_{(R*[[e_0|k]])} e''$. Note that:

(iv) $e_0 | k = e' | k$,

(v) $e(k) = e'(k) \neq e'' | k$, and

(vi) $k+1 | e'' = k+1 | e'$. By (ii, iv),

(vii) $e' \notin \min((R * [[e' | k]]), [k, e'(k)])$. By (v),

(viii) $e'' \notin \min((R * [[e' | k]]), [k, e'(k)])$. So by (iii, v, vii, viii) and clause (2) of the definition of $*_M$:

(ix) $e' \geq_{(R*[[e'|k+1]])} e''$. So by (vi) and positive order-invariance (proposition 7), for all $k' \geq k+1$, $e' \geq_{(R*[[e'|k']])} e''$; contradicting the reductio hypothesis. ⊣

Definition 18 *e is propped up at n in r just in case for each $e' \in [e]^\leq_r$, $e'(n) \neq e(n)$. e is propped up in r just in case there exists an n such that e is propped up at n in r.*

28

Proposition 22 (propping condition for $*_M$) *If $(r, *_M)$ identifies $dom(r)$ then for each $e \in dom(r)$, for each m, there is an $m' \geq m$ such that e is propped up in $(r *_M [[e|m]])$ at m'; so in particular, e is propped up in r.*

Proof: Suppose that for all $m' \geq m$, $e \in dom(r *_M [[e|m]])$ is not propped up in $(r *_M [[e|m]])$ at m'. Using the definition of $*_M$, show by a straightforward induction on $k - m$ that for all $k \geq m, k' \geq k$, e is not propped up in $(r *_M [[e|k]])$ at k'. Hence, for all $k \geq m$, $(r *_M [[e|k]])(e) > 0$, so $(r, *_M)$ does not identify $dom(r)$. $\dashv$

The following definition generalizes the notion of isolated points to the case in which there is sufficient information after a given position n to distinguish e from all other points in S. k-isolation is more stringent than isolation when $k > 0$. For example, 0^∞ is isolated but not 1-isolated in $\{10^n 1^\infty : n \in \omega\}$.

Definition 19 *e is k-isolated in $S \Leftrightarrow$ there exists an $n \geq k$ such that $[(k|e)|n] \cap S \subseteq \{e\}$.*

Proposition 23 (isolation condition) *If $*$ is positively order-invariant and $(r, *)$ identifies $dom(r)$ then for each $e \in dom(r)$, for all k, e is k-isolated in $[e]^{\leq}_{(r*[[e|k]])}$; so in particular, e is isolated in $[e]^{\leq}_r$.*

Proof: Suppose e is not k-isolated in $[e]^{\leq}_{(r*[[e|k]])}$. Then for each $n \geq k$, there is an $e_n \neq e$ such that $r(e_n) \leq r(e)$ and $(k|e_n)|n = (k|e)|n$. So by positive order-invariance, for each n, if $(r * [[e|n]])(e) = 0$ then $(r * [[e_n|n]]) = 0$. Hence, $(r, *)$ does not identify $dom(r)$. $\dashv$

Proposition 24 *If $*$ is positively order-invariant and $(r, *)$ identifies $G^{n+1}(e_0) \subseteq dom(r)$ and $e \in G^n(e_0)$ then for all k, for all but finitely many j, $(r * [[e_0|k]])(e) < (r * [[e_0|k]])(e \ddagger j)$.*

Proof: Let $e \in G^n(e_0)$. Then for each j, $e \ddagger j \in dom(r)$. Suppose that for some k there are infinitely many distinct j such that $(r * [[e|k]])(e) \geq (r * [[e|k]])(e \ddagger j)$. Then e is not k-isolated in $[e]^{\leq}_{(r*[[e_0|k]])}$. So by proposition 23 and the fact that $*$ is positively order- invariant, $(r, *)$ does not identify $G^{n+1}(e_0)$. $\dashv$

Proposition 25 (stacking lemma) *For all $k, n, n' \leq n$, if $*$ is positively order-invariant and $(r, *)$ identifies $G^n(e_0) \subseteq dom(r)$ and $(r * [[e_0|k]])(e_0) = 0$ then there exists an $e_{n'}$ such that*

1. *$e_{n'} \in g^{n'}(e_0)$,*

2. *$e_0|(k+1) = e_{n'}|(k+1)$ and*

3. *$(r * [[e_0|k]])(e_{n'}) \geq n'$.*

Proof: Assume the antecedent. Let n, k be given. We show the consequent by induction on $n' \leq n$. When $n' = 0$, (1-3) are trivially satisfied by e_0. Now suppose that $n' + 1 \leq n$ and that there exist $e_0, \ldots e_{n'}$ satisfying (1-3). Since $n' + 1 \leq n$, $(r, *)$ identifies $G^{n'+1}(e_0)$. So by proposition 24, we may choose j sufficiently large so that so that
(i) $(r * [[e_0|k]])(e_{n'}) < (r * [[e_0|k]])(e_{n'} \ddagger j)$ and
(ii) $j > \max(\Gamma(e_0, e_{n'}) \cup \{k+1\})$. Now set $e_{n'+1} = e_{n'} \ddagger j$. $e_{n'+1}$ satisfies (1, 2) because of (ii) and the fact that $e_{n'}$ does. $e_{n'+1}$ satisfies (3) because of (i) and the fact that $e_{n'}$ does. $\dashv$

Proposition 26 *If $*$ is positively order-invariant and $(r, *)$ identifies $G^{n+1}_{even}(e_0) \subseteq dom(r)$ and $e \in G^n_{even}](e_0)$ then for all k, for all but finitely many j, for all $m > 0$, $(r * [[e_0|k]])(e) < (r * [[e_0|k]])((e' \ddagger j) \ddagger j + m)$.*

Proof: Similar to the proof of proposition 24. $\dashv$

Proposition 27 (even stacking lemma) *Proposition 25 continues to hold when G^n, g^n are replaced with G^n_{even}, g^n_{even}.*

29

Proof: Similar to the proof of proposition 25, using proposition 26. ⊣

Proposition 28 (with Oliver Schulte) *For all $e_0 \in B$, For all $j \geq 2$, $G^j(e_0)$ is identifiable using just j retractions, but is not identifiable by $*_{DP,1}$.*

Proof: The positive claim is from proposition 11. For the negative claim, suppose for reductio that there is an IA r such that $(r, *)$ identifies $G^j(e_0)$, where $* = *_{DP,1}$ and $j \geq 2$. Then since $e_0 \in G^0(e_0)$ and $0 < j$, there exists a least n such that

(i) $(r, *)(e_0|n) = \{e_0\}$. Then there exists a least $k > n$ such that

(ii) $(r, *)((e_0 \ddagger n)|k) = \{e_0 \ddagger n\}$, since $e_0 \ddagger n \in G^1(e_0)$ and $1 < j$. Define $R = \{e'' \in U : |\Gamma(e_0, e'')| \text{ is odd}\}$. Since $|\Gamma(e_0, (e_0 \ddagger n))| = 1$ is odd, we have by (i) and (ii) that there is a least $k' > n$ such that

(iii) $b(r * [[(e_0 \ddagger n)|k']])(e) \cap G^j(e_0) \subseteq R$. Since k' is least, there exists an e such that

(iv.a) $(r * [[(e_0 \ddagger n)|(k' - 1)]])(e) = 0$,

(iv.b) $(r * [[(e_0 \ddagger n)|k']])(e) > 0$, and

(iv.c) $e \in G^j(e_0) - R$. Since $* = *_{DP,1}$, we also have[29]

(iv.d) $(r * [[(e_0 \ddagger n)|k']])(e) = 1$.

Case 1: $k' = n + 1$. Then $e = e_0$, by (i), (iv.a). Define $e' = (e_0 \ddagger n) \ddagger n + 1$. Hence,

(v.a) $e'|n + 1 = (e_0 \ddagger n)|n + 1$,

(v.b) $n + 1|e' = n + 1|e_0$, and

(v.c) $e' \in G^j(e_0) - R$, since $|\Gamma(e_0, e')| = 2$ and $j \geq 2$.[30] By (v.a, c) and the reductio hypothesis, $e' \in \mathrm{dom}(r*[[(e_0\ddagger n)|n+1]])$, else $(r, *)$ fails to identify $e' \in G^j(e_0)$. So by (iii), (v.a,c), and the case hypothesis,

(vi) $(r * [[e'|n + 1]])(e') \geq 1$. By (iv.d), (v.a), and the case hypothesis,

(vii) $(r * [[e'|n + 1]])(e) = 1$. By positive order invariance (proposition 7) and (v.b), (vi), (vii), we have that for each $m \geq n + 1$, $(r * [[e'|m]])(e') \geq (r * [[e'|m]])(e_0)$, contradicting the reductio hypothesis.

Case 2: $k' \geq n + 2$. Then by the definition of $*$ and (iv.a,b), we have

(viii.a) $e(k' - 2) = (e_0 \ddagger n)(k' - 2) = \neg e_0(k' - 2)$ and

(viii.b) $e(k' - 1) = \neg(e_0 \ddagger n)(k' - 1) = e_0(k' - 1)$. Let e' be defined so that:

(ix.a) $e'|k' = (e_0 \ddagger n)|k'$, and

(ix.b) $k'|e' = k'|e$. By (viii.a), there exists some $j < k' - 2$ such that $j \in \Gamma(e_0, e)$. By (viii.a,b), $k'-1 \in \Gamma(e_0, e)$. So $|\{j \leq k' : j \in \Gamma(e_0, e)\}| \geq 2$. But by (ix.a,b) we also have $|\{j \leq k' : j \in \Gamma(e_0, e')\}| \leq 2$. So by (ix.b) and (iv.c),

(x) $e' \in G^j(e_0)$. So by the reductio hypothesis and (ix.b),

(xi) $e' \in \mathrm{dom}(r * [[(e \ddagger n)|k']])$, else $(r, *)$ does not identify $e' \in G^j_{\mathrm{even}}(e_0)$. By (iv.c), $|\Gamma(e_0, e)|$ is even. Hence, e agrees almost everywhere with e_0. By (ix.b), e' agrees almost everywhere with e and hence with e_0. So, $|\Gamma(e_0, e')|$ is even. So $e' \notin R$. Thus, by (iii, xi), $(r * [[(e_0 \ddagger n)|k']])(e') \geq 1$. So by positive invariance, (iv.d, ix.a,b), we have that for all $m \geq k'$, $(r * [[e'|m]])(e') \geq (r * [[e'|m]])(e_0)$, contradicting the reductio hypothesis. ⊣

Proposition 29 *Let $e_0 \in B$.*

1. $G^1(e_0)$ is identifiable by $*_{DP,1}$.

2. For all $j \geq 1$, $G^j(e_0)$ is not identifiable by $*_M, *_{S,1}, *_{A,1}$.

Proof of (1): Let $* = *_{DP,1}$. Define $r^{-1}(0) = g^0(e_0) = \{e_0\}$ and $r^{-1}(1) = g^1(e_0) = \{e_0 \ddagger k : k \in \omega\}$. Then $\mathrm{dom}(r) = G^1(e_0)$. Let $e \in G^1(e_0)$. Case: $e = e_0$. Then by timidity and stubbornness, we have that for each $k, b(r * [[e|k]]) = \{e\}$. Case: for some n, $e = e_0 \ddagger n$. Suppose $k \leq n$. On data $e|k$, e_0 stays uniquely

---

[29]This is where the value $\alpha = 1$ enters the negative argument.

[30]This is where $j \geq 2$ enters the argument.

at level 0 by timidity and stubbornness. Positive invariance keeps e at level 1, along with all the $e_0 \ddagger n'$ such that $n' \geq k$. For each $n' < k$, the climbing lemma implies that $e_0 \ddagger n'$ is at least as high as level $k - n'$ on data $e|k$. So by the time $e|n$ has been read, we have

(i.a) $b(r * [[e|n]]) = \{e_0\}$,

(i.b) if $n' \geq n$ then $(r * [[e|n]])(e_0 \ddagger n') = 1$, and

(i.c) if $n' < n$ then $(r * [[e|n]])(e_0 \ddagger n') > 1$. On data $e|n + 1$, e_0 is refuted and moves up one level along with all data streams of form $e_0 \ddagger n'$, where $n' > n$. By (i.a,b,c), e is the lowest data stream consistent with the data, so e drops to level 0. All data streams of form $e \ddagger n'$ such that $n' < n$ also drop one level with e, but fortunately, by (i.c) they all end up above level 0. So $b(r * [[e|n + 1]]) = \{e\}$. By timidity and stubbornness, e remains uniquely at level 0 forever after.

Proof of (2): Case: $* = *_{A,1}$. Instance of proposition 30. Case: $* = *_{S,1}, *_M$. Suppose for reductio that there is an IA r such that $(r, *)$ identifies $G^1(e_0)$. Then there exists a least n such that

(i) $b(r * [[e_0|n]]) = \{e_0\}$. Furthermore,

(ii) $\exists k \geq n$ such that $\forall k' > k$, $(r * [[e_0|n]])(e_0 \ddagger k') \geq (r * [[e_0|n]])(e_0 \ddagger k)$; for otherwise, there would exist an infinite descending chain of ordinals in the range of $(r * [[e_0|n]])$. By (ii), there exists a $k \geq n$ such that

(iii) $(r * [[e_0|n]])(e_0 \ddagger k + 1) \geq (r * [[e_0|n]])(e_0 \ddagger k)$. Observe that:

(iv) $(e_0 \ddagger k)|k = (e_0 \ddagger k + 1)|k = e_0|k$ and

(v) $(e_0 \ddagger k + 1)(k) = e_0(k) \neq (e_0 \ddagger k)(k)$. By timidity and stubbornness (proposition 7) and (i, iv, v),

(vi) $\forall n', n \leq n' \leq k + 1 \Rightarrow b(r * [[(e_0 \ddagger k + 1)|n']]) = \{e_0\}$. By (iii, iv, vi) and positive order-invariance (proposition 7),

(vii) $\forall n', n \leq n' \leq k \Rightarrow (r * [[(e_0 \ddagger k + 1)|n']])(e_0 \ddagger k + 1) \geq (r * [[(e_0 \ddagger k + 1)|n']])(e_0 \ddagger k) > 0$. Now it is claimed as well that:

(viii) $(r * [[(e_0 \ddagger k + 1)|k + 1]])(e_0 \ddagger k + 1) \geq (r * [[(e_0 \ddagger k + 1)|k + 1]])(e_0 \ddagger k)$. For consider the case of $*_M$. By (v, vi) and the definition of $*_M$,

$(r *_M [[(e_0 \ddagger k + 1)|k + 1]])(e_0 \ddagger k) = (r *_M [[(e_0 \ddagger k + 1)|k]])(e_0 \ddagger k) + 1$ and

$(r *_M [[(e_0 \ddagger k + 1)|k + 1]])(e_0 \ddagger k + 1) = (r *_M [[(e_0 \ddagger k + 1)|k]])(e_0 \ddagger k + 1) + 1$. So by (vii), we have (viii) for $*_M$.

Let us turn now to the case of $*_{S,1}$. By (v, vi),

$\min\{(r * [[(e_0 \ddagger k + 1)|k]])(e') : e' \in dom(r * [[(e_0 \ddagger k + 1)|k]]) \cap [k, (e_0 \ddagger k + 1)(k)]\} = 0$ and

$\min\{(r * [[(e_0 \ddagger k + 1)|k]])(e') : e' \in dom(r * [[(e_0 \ddagger k + 1)|k]]) \cap (U - [k, (e_0 \ddagger k + 1)(k)])\} > 0$. So by (v) and the definition of $*_{S,1}$,

$(r *_{S,1} [[(e_0 \ddagger k + 1)|k + 1]])(e_0 \ddagger k + 1) = -0 + (r *_{S,1} [[(e_0 \ddagger k + 1)|k]])(e_0 \ddagger k + 1)$ and

$(r *_{S,1} [[(e_0 \ddagger k + 1)|k + 1]])(e_0 \ddagger k) \geq -1 + (r *_{S,1} [[(e_0 \ddagger k + 1)|k]])(e_0 \ddagger k)$. So again by (vii) we have (viii) for $*_{S,1}$.

Finally, since $(k + 1)|(e_0 \ddagger k) = (k + 1)|(e_0 \ddagger k + 1)$, we have by (viii) and positive order-invariance that for all $k' \geq k + 1$,

$(r * [[(e_0 \ddagger k + 1)|k']])(e_0 \ddagger k + 1) \geq (r * [[(e_0 \ddagger k + 1)|k']])(e_0 \ddagger k)$, contradicting the reductio hypothesis. ⊣

Proposition 30 (restrictiveness of $*_{A,n}$) *Let $e_0 \in B$.*

1. $G^0(e_0)$ is identifiable by $*_{A,0}$.

2. for all n, $G^{n+1}(e_0)$ is identifiable by $*_{A,n+2}$.

3. forall $m \geq n$, $G^{m+1}(e_0)$ is not identifiable by $*_{A,n+2}$. $*_{A,n+1}$.

Proof of (1): Let $dom(r) = \{e_0\}$ and let $r(e_0) = 0$. Then for all k, $(r, *_{S,0})(e_0|k) = \{e_0\}$.

Proof of (2): by propositions 9 and 11.

Proof of (3): Suppose for reductio that $(r, *_{A,n+1})$ identifies $G^{m+1}(e_0)$, with $m \geq n$. Then for some j,

(i) $b(r *_{A,n+1} [[e_0|j]]) = \{e_0\}$. So by positive invariance (proposition 7) and $G^{m+1}(e_0) \subset dom(r)$ by the

reductio hypothesis, proposition 25 yields

(ii) there exists an $e \in G^{n+1}(e_0) - G^n(e_0)$ such that $e_0|j+1 = e|j+1$ and $(r * [[e_0|j]])(e) \geq n+1$. $e \neq e_0$, so let z be least such that $e(z) \neq e_0(z)$. So,

(iii) $z > j$. So since $z > 0$, we may define e' to be just like e except that $e'(z-1) = \neg(e_0(z-1))$. Hence,

(iv) $e'(z-1) \neq e(z-1) = e_0(z-1)$. Also

(v) $e|z = e_0|z$ and

(vi) $z|e' = z|e$. By (i, v), and the timidity of $*_{A,n+1}$ (proposition 7),

(vii) for all x such that $j \leq x \leq z$, $(r *_{A,n+1} [[e|x]])(e_0) = 0$. By positive invariance and (ii, v, vii),

(viii) $(r *_{A,n+1} [[e|z]])(e) \geq n+1$. By (iv) and the definition of $*_{A,n+1}$,

(ix) $(r *_{A,n+1} [[e|z]])(e) = n+1$. By (vi, viii, ix) and positive invariance,

(ix) for all $k' \geq z, (r *_{A,n+1} [[e|k']])(e) = (r *_{A,n+1} [[e|k']])(e')$. Hence, $(r, *_{A,n+1})$ does not identify $G^{n+1}(e_0)$, contradicting the reductio hypothesis. ⊣

Proposition 31 *Let $e_0 \in B$.*

1. *$G^\omega_{even}(e_0)$ is identifiable by $*_{S,1}, *_{DP,1}$.*

2. *$\forall m \geq 1, G^m_{even}(e_0)$ is not identifiable by $*_M$.*

3. *$G^0_{even}(e_0)$ is identifiable by $*_{A,0}$*

4. *$G^n_{even}(e_0)$ is identifiable by $*_{A,n+1}$*

5. *$\forall m \geq n, G^{m+1}_{even}(e_0)$ is not identifiable by $*_{A,n+1}$*

Proof of (1). Case: $* = *_{DP,1}$. Let r_e be r_e^H restricted to $G^\omega_{even}(e)$, so for each $e' \in G^\omega_{even}(e)$, $r_e(e') = \rho(e, e')$. Let e be given and let $e' \in G^\omega_{even}(e)$. Define e'_i so that

(i.a) $e'_i|i = e'|i$ and (i.b) $i|e'_i = i|e$. Hence, $e = e'_0$. Also let

$c(e', k) = 1+$ the greatest $i < k$ such that $e'(i) \neq e(i)$, if there is such an i, and let $c(e', k) = 0$ otherwise. Recall that $G^\omega_{even}(e)$ is precisely the set of all finite variants of e, so $\Delta(e, e')$ is finite. Let $m(e') = 1+ \max(\Delta(e, e'))$. Then

(ii) for all $k \geq m(e'), e_{c(e',k)} = e'$. I claim that $*$ satisfies the following symmetry condition: for each $e', e'' \in G^\omega_{even}(e)$,

(iii.a) $(r_e * [[e'|k]])(e'') \geq r_{e'_{c(e',k)}}(e'')$, and

(iii.b) if $e''|k = e'|k$ then $(r_e * [[e'|k]])(e'') = r_{e'_{c(e',k)}}(e'')$. Then for each $e' \in G^\omega_{even}(e)$, for each $k \geq m(e')$, $b(r_e * [[e'|k]]) = b(r_{e'_{c(e',k)}}) = \{e'_{c(e',k)}\} = \{e'\}$ (by ii). Thus, $(r_e, *)$ identifies $G^\omega_{even}(e)$. So it remains only to establish (iii.a, b). (iii.a, b) are immediate when $k = 0$. Now suppose (iii.a,b) hold at k. Then

(iv) $b(r_e * [[e'|k]]) = \{e'_{c(e',k)}\}$. Let $var(e'_{c(e',k)})$ be just like $e'_{c(e',k)}$ except that $var(e'_{c(e',k)})(k) = \neg(e'_{c(e',k)})(k)$. $\rho(var(e'_{c(e',k)}), e'_{c(e',k)}) = 1$ and $e'|k = var(e'_{c(e',k)})|k$ so (iii.b) of the induction hypothesis yields

(v) $(r_e * [[e'|k]])(var(e'_{c(e',k)})) = 1$.

Case 1: $c(e', k+1) = k+1$. Then $var(e'_{c(e',k)})(k) = e'(k) \neq e'_{c(e',k)}(k)$. So by (iv, v),

(vi) $\min\{(r * [[e'|k]])(e'') : e'' \in dom(r * [[e'|k]]) \cap [k, e'(k)]\} = 1$. Now let $e'' \in G^\omega_{even}(e)$. Subcase: $e''(k) = e'(k)$. Then by the induction hypothesis and (vi),

$(r_e * [[e'|k+1]])(e'') = -1+(r_e * [[e'|k]])(e'') \geq -1+\rho(e_{c(e',k)}, e'') = \rho(e_{c(e',k+1)}, e'') = r_{e_{c(e',k+1)}}(e'')$. When $e''|k+1 = e'|k+1$, the inequality is strenghened to an equality by (iii.b) of the induction hypothesis, yielding

$(r_e * [[e'|k]])(e'') = r_{e_{c(e',k+1)}}(e'')$. Subcase: $e''(k) \neq e'(k)$. Then by the induction hypothesis,

$(r_e * [[e'|k+1]])(e'') = (r_e * [[e'|k]])(e'') + 1 \geq \rho(e_{c(e',k)}, e'') + 1 = \rho(e_{c(e',k+1)}, e'') = r_{e_{c(e',k+1)}}(e'')$. Since $e''(k) \neq e'(k)$, $e'|k+1 \neq e''|k+1$ so (iii.b) is trivial in this subcase.

Case 2: $c(e', k+1) < k+1$. Then

(vii) $e_{c(e',k)} = e_{c(e',k+1)}$. Hence, $var(e'_{c(e',k)})(k) \neq e'(k) = e'_{c(e',k)}(k)$. So by (iv),

32

(viii) $\min\{(r * [[e'|k]])(e'') : e'' \in \text{dom}(r * [[e'|k]]) \cap [k, e'(k)]\} = 0$. Now let $e'' \in G^\omega_{\text{even}}(e)$. Subcase: $e''(k) = e'(k)$. Then by the induction hypothesis and (vii, viii),

$(r_e * [[e'|k+1]])(e'') = -0 + (r_e * [[e'|k]])(e'') \geq \rho(e_{c(e',k)}, e'') = \rho(e_{c(e',k+1)}, e'') = r_{e_{c(e',k+1)}}(e'')$. When $e''|k+1 = e'|k+1$, the inequality is strengthened to an equality by (iii.b) of the induction hypothesis, yielding

$(r_e * [[e'|k]])(e'') = r_{e_{c(e',k+1)}}(e'')$. Subcase: $e''(k) \neq e'(k)$. Then by the induction hypothesis and (vii),

$(r_e * [[e'|k+1]])(e'') = (r_e * [[e'|k]])(e'') + 1 \geq \rho(e_{c(e',k)}, e'') = \rho(e_{c(e',k+1)}, e'') = r_{e_{c(e',k+1)}}(e'')$. Since $e''(k) \neq e'(k)$, $e'|k+1 \neq e''|k+1$ so (iii.b) is trivial in this subcase.

Case $* = *_{S,1}$. The argument is similar to the preceding one, except that the symmetry condition (iii.a,b) can be strengthened to:

(iii) for each $e' \in G^\omega_{\text{even}}(e)$, $(r_e * [[e'|k]]) = r_{e'_{c(e',k)}}$, which implies the success of $(r, *)$ as before.[31] (iii) is immediate when $k = 0$.

In case 1, the induction hypothesis yields (vi) as well as

(vi') $\min\{(r * [[e'|k]])(e'') : e'' \in \text{dom}(r * [[e'|k]]) \cap U - [k, e'(k)]\} = 0$. Subcase $e''(k) = e'(k)$, is as before, with an equality replacing the inequality. In subcase $e''(k) \neq e'(k)$, (vi, vi') yield:

$(r_e * [[e'|k+1]])(e'') = -0 + (r_e * [[e'|k]])(e'') + 1 = \rho(e_{c(e',k)}, e'') + 1 = \rho(e_{c(e',k+1)}, e'') = r_{e_{c(e',k+1)}}(e'')$.

In case 2, the induction hypothesis yields (viii) as well as

(viii') $\min\{(r * [[e'|k]])(e'') : e'' \in \text{dom}(r * [[e'|k]]) \cap U - [k, e'(k)]\} = 1$. Subcase $e''(k) = e'(k)$ is as before, with an equality replacing the inequality. In subcase $e''(k) \neq e'(k)$, (viii, viii') yield:

$(r_e * [[e'|k+1]])(e'') = -1 + (r_e * [[e'|k]])(e'') + 1 = \rho(e_{c(e',k)}, e'') = \rho(e_{c(e',k+1)}, e'') = r_{e_{c(e',k+1)}}(e'')$.

Proof of (2): Let $* = *_M$. Suppose for reductio that there is an IA r such that $(r, *)$ identifies $G^1_{\text{even}}(e_0)$. Then there exists a least n such that

(i) $(r, *)(e_0|n) = (r, *)(e_0) = \{e_0\}$. Furthermore,

(ii) $\exists i \geq n, (r * [[e_0|n]])((e_0 \ddagger i) \ddagger i + 2) \leq (r * [[e_0|n]])((e_0 \ddagger i + 1) \ddagger i + 3)$, else, there would exist an infinite descending chain of ordinals in the range of $(r * [[e_0|n]])$. Let $e = (e_0 \ddagger i) \ddagger i + 2$, $e' = (e_0 \ddagger i + 1) \ddagger i + 3$, and $e'' = (e_0 \ddagger i) \ddagger i + 3$.

Case 1: $(r * [[e_0|n]])(e'') > (r * [[e_0|n]])(e')$. Then by (i,ii), e'' is not propped up in $(r * [[e_0|n]])$, contradicting the reductio hypothesis by proposition 22.

Case 2: $(r * [[e_0|n]])(e'') \leq (r * [[e_0|n]])(e')$. $e_0|i = e'|i = e''|i$, so by timidity and positive order-invariance (proposition 7),

(iii) $(r * [[e'|i]])(e_0) = 0 < (r * [[e'|i]])(e'') \leq (r * [[e'|i]])(e')$. So by timidity, stubbornness, (iii) and the fact that $e'|i+1 = e_0|i+1$, we have:

(iv) $b(r * [[e'|i+1]])(e_0) = \{e_0\}$. So $e', e'' \notin b(r * [[e'|i+1]])(.|[i, e'(i)])$. So by the definition of $*_M$ and (iii), $(r*[[e'|i+1]])(e'') = (r*[[e'|i]])(e'')+1 \leq (r*[[e'|i]])(e')+1 = (r*[[e'|i+1]])(e')$. So since $i+1|e' = i+1|e''$, positive order-invariance yields that for all $k \geq i+1$, $(r * [[e'|k]])(e'') \leq (r * [[e'|k]])(e')$, contradicting the reductio hypothesis.

Proof of (3): Immdediate.

Proof of (4): Immediate consequence of propositions 9 and 11.

Proof of (5): The argument is identical to the one provided for proposition 30 except that the appeal to proposition 25 is replaced with an appeal to proposition 27. ⊣

---

[31] Condition (iii) implies the hypercube rotation representation of the evolution of $(*, r)$, as was mentioned in the informal discussion of this proposition.

33