The Diachronic Coherence of Ungraded Beliefs

Matthias Hild
Christ's College, Cambridge

5 November, 1998

Abstract. This paper works within a model of ungraded belief that characterizes epistemic states as logically closed and consistent sets of sentences. The aim of this paper is to discuss three diachronic coherence conditions for such beliefs. These coherence conditions are formulated in terms of the reasoner's present beliefs about how his present beliefs will evolve in the future, for instance, in response to different pieces of future evidence.

Key words: Belief Revision, Coherence, Surprise Examination.

Introduction

This paper assumes that an epistemic state, or at least some important aspect of it, can be represented by a set of ungraded beliefs and, more specifically, by a logically closed and consistent set of sentences.[1]

Section 1 asks in what conceptual context the notion of ungraded belief is situated. In my view, the absence of a satisfactory answer overshadows the entire field and, in particular, the discussion in this paper.

Section 2 introduces an auto–epistemic reasoner who has the capacity to reason at any point in time about the beliefs he holds at that point in time. I call such reasoning 'synchronic'. Synchronic auto–epistemic reasoning has first been discussed in terms of the synchronic iteration properties of the knowledge and belief operator (Hintikka (1962)). Synchronic standards of auto–epistemic coherence have subsequently been related to default reasoning (Konolige (1987)).

Section 3 pursues a parallel programme in the diachronic context in which the reasoner has the capacity to reason at any point in time about his beliefs at other points in time. He may, in particular, reason about how he will change his beliefs in the future, e.g., upon receiving posterior evidence. I discuss three diachronic coherence constraints on the reasoner's prior beliefs about his posterior beliefs.

Section 4 examines how coherent auto–epistemic reasoning about posterior evidence affects learning from evidence. I argue that methods or theories about how to learn from evidence reduce to prior beliefs about the different pieces of posterior evidence that the reasoner might receive. Rules for learning from evidence, including the Alchourrón/Gärdenfors/Makinson (AGM) theory of belief revision thus interpreted, are coherent (only) in special cases where the reasoner holds suitable prior beliefs about his learning situation.

1. Ungraded Belief

What is it?

The strongest defense of any (normative) coherence principle is, I think, one that relates different types of coherence concepts with each other. Bayesian epistemology, for instance, has capitalized on drawing a very close connection between decision–theoretic and probabilistic coherence. The famous Dutch Book argument shows that if you wish to be decision–theoretically coherent (and to maximize expected utility), then you have to conform to the probability calculus and *vice versa*.

We might perhaps wish to emulate this Bayesian line of defense and relate coherence conditions for ungraded beliefs to other types of coherence concepts. But what other coherence concepts? What is the conceptual context of the notion of ungraded beliefs? What is the function — cognitive, decision–theoretic, or otherwise — of ungraded beliefs? At present, I can only mention some attempts at answering such questions:

1. Ungraded belief in A is a disposition to assent to the question whether A is true when asked under suitable circumstances.

2. Ungraded belief in A is a disposition to act as if A were true relative to any (alternatively, some) practical objective.

3. Ungraded belief is a proposition that has a personal probability above some value $t \in [0.5, 1]$.

4. The adoption of an ungraded belief is a decision under uncertainty that maximizes the P–expectation of the reasoner's cognitive utility relative to his personal probabilities P.

5. The structure of ungraded beliefs is a crude, qualitative version of the quantitative Bayesian model.

Suggestions 2, 3 and 5 try to furnish a welcome connection between ungraded beliefs and action. If, for instance, we define an ungraded belief as a proposition with personal probability $p = 1$ (the '$p = 1$ definition'), then the coherence principles in this paper become special cases of Bayesian principles that are supported by a decision–theoretic coherence condition (Hild (1998a)(1998b)). I intend to discuss some serious problems with Suggestions 1–5 elsewhere. For the time being, I cannot but presuppose some preliminary understanding of ungraded belief based mainly on Suggestion 1. In my discussion of coherence principles, this reduces me to appealing to some preliminary fragments of a future theory, i.e., to your intuition.

Let $\mathcal{L}$ be a propositional language that is closed under the classical truth–functional connectives $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$. Let $\mathcal{S}$ be the set of all subsets of $\mathcal{L}$ that are consistent and closed with respect to classical propositional logic. Any $K \in \mathcal{S}$ is called a *belief set*. We assume in this paper that a reasoner's ungraded beliefs are represented by some belief set K.

Conditional belief

If a reasoner's ungraded beliefs are represented by some belief set K and if $A \in \mathcal{L}$, then $K|A$ denotes the reasoner's belief set conditional on A. Again, various attempts have been made at embedding conditional beliefs in some wider theory (whilst the problem of interpreting unconditional beliefs persists):

1. Conditional ungraded beliefs are connected to unconditional ungraded beliefs in (certain types of) conditionals (e.g., via the Ramsey Test).

2. Conditional ungraded beliefs are the result of hypothetical reasoning or supposing (Rescher (1964)).

3. The ungraded beliefs conditional on A are the ungraded beliefs after having learnt A.

Problems with Suggestion 1 are much discussed in the literature. Suggestion 2 seems little informative since suppositional reasoning is as much in need of an explication as conditional beliefs. Suggestion 3 is untenable, or so I will argue in Section 4.

Applied to conditional beliefs, the AGM theory of 'belief revision' (Gärdenfors (1988)) proposes the following axioms (for all $A, B \in \mathcal{L}$ and $K \in \mathcal{S}$):[2]

$(K|1)$ $K|A$ is closed under classical propositional logic.

$(K|2)$ $A \in K|A$, unless $A \Leftrightarrow \bot$.

$(K|3)$ $K|A \subseteq K^{+}A$.

$(K|4)$ If $\neg A \notin K$, then $K^{+}A \subseteq K|A$.

($K|5$) $\perp \notin K|A$, unless $A \Leftrightarrow \perp$.

($K|6$) If $A \Leftrightarrow B$, then $K|A = K|B$.

($K|7$) $K|(A \wedge B) \subseteq (K|A)^+B$.

($K|8$) If $\neg B \notin K|A$, then $(K|A)^+B \subseteq K|(A \wedge B)$.

I will now propose an alternative characterization of (some of) the properties of conditional beliefs that is motivated by Suggestion 1. It recommends itself by presupposing only a weak connection between conditional beliefs and truth–functional conditionals. My axiomatization turns out to be equivalent to the AGM axioms ($K|1$)–($K|8$) which are normally motivated as a theory of conservative, minimal belief change and are thus typically based on Suggestion 3 and perhaps Suggestion 2.[3] Let the *Logical Core* consist of Closure ($K|1$), Consistency ($K|5$), Extensionality ($K|6$) and Reduction.

(Reduction) $K|\top = K$

I propose the following central axiom: A truth–functional conditional $A \rightarrow B$ is believed if and only if it is either vacuously believed or B is believed conditionally on A (for all $A, B \in \mathcal{L}$):[4]

(Weak Multiplication) $(A \rightarrow B) \in K$ iff either $\neg A \in K$, or $B \in K|A$, or both.

Strong Multiplication takes a step towards the treatment of nested conditionals (for all $A, B, C \in \mathcal{L}$):

(Strong Multiplication) $(A \rightarrow B) \in K|C$ iff either $\neg A \in K|C$, or $B \in K|(A \wedge C)$, or both.

Lindström/Rabinowicz (this volume) have independently proposed the same axiom. Strong Multiplication is perhaps less evident than Weak Multiplication, but notice that it follows from Weak Multiplication if we make the following assumption about nested conditional beliefs (for all $A, C \in \mathcal{L}$):

(N) If $\neg A \notin K|C$, then $K|C|A = K|(A \wedge C)$.

(N), in turn, follows from Strong Multiplication in the presence of ($K|1$), ($K|6$), and Reduction.[5] In total, I suggest the following axiom system for conditional, ungraded, belief: Logical Core, Success ($K|2$) and at least Weak Multiplication.

Theorem 1.1. Given $(K|1)$ and $(K|6)$, Weak Multiplication plus Reduction is equivalent to $(K|3)$ plus $(K|4)$.

Given $(K|1)$ and $(K|6)$, Strong Multiplication plus Reduction is equivalent to $(K|3)$, $(K|4)$, $(K|7)$, plus $(K|8)$.

For the proof see Hild (1998b).

2. Synchronic Auto–Epistemology

Before I turn to diachronic coherence conditions, let me briefly review some important concepts in the theory of synchronic coherence. We proceed in three steps. First, we enrich the reasoner's language with auto–epistemic vocabulary. Second, we impose a synchronic coherence condition on auto–epistemic beliefs. Finally, we apply auto–epistemic coherence in an analysis of default reasoning. In the next section I will treat the diachronic case in an exactly parallel fashion.

Step 1. Let $\mathcal{L}^0$ be a propositional language that is closed under the classical truth–functional connectives. Starting from this basic language, we construct its auto–epistemic extension $\mathcal{L}^{AE}$ that is closed under the classical truth–functional connectives as well as under the sentential operator $\mathbf{B}(.)$ (i.e., $\mathbf{B}(A) \in \mathcal{L}^{AE}$, for all $A \in \mathcal{L}^{AE}$). We take the sentence $\mathbf{B}(A) \in \mathcal{L}^{AE}$ to express the proposition that the reasoner believes A. Such propositions about the reasoner's beliefs can themselves be part of what the reasoner believes. Thus the reasoner can form auto–epistemic beliefs, i.e., beliefs about her own beliefs. Let K be a variable ranging over the set of all $\mathcal{L}^{AE}$–belief sets (i.e., the set of all subsets of $\mathcal{L}^{AE}$ that are consistent and closed with respect to classical propositional logic).

Step 2. Stalnaker (1980) proposed the following synchronic coherence condition (under the title of a 'stable set'):

(AE–Transparency) $\quad A \in K$ iff $\mathbf{B}(A) \in K$
$\qquad\qquad\qquad\qquad A \notin K$ iff $\neg\mathbf{B}(A) \in K$

This condition demands perfect Cartesian introspection of one's own beliefs. It implies, for instance:

(Anti–Moore) $\quad (\mathbf{B}(A) \wedge \neg A) \notin K$.
Don't *assume* your beliefs to be wrong!

Short of implying cognitive infallibility, AE–Transparency admits the possibility that I believe A while A is in fact false. Instead, it requires me not to undermine my own epistemic state by believing that some of my beliefs are false. As long as I have undermining doubts about my belief in A, I should not be satisfied. In a state of ideal belief, doubts about the truth of my own beliefs must be dissolved.

Step 3. A default is an inference rule that forces its conclusion upon a reasoner on the basis of two sets of conditions, firstly, the reasoner's belief in some (conjunction of) supporting premises and, secondly, the absence of the reasoner's belief in some contravening premises (cf. Reiter (1980)). We say that a belief set is closed under default rule (D) when the following holds:

$$\text{(D)} \qquad \text{If } A \in K \text{ and } B_1, \ldots, B_N \notin K, \text{ then } C \in K.$$

Assuming that auto–epistemic beliefs are synchronically coherent (AE–Transparency), any default (D) can be translated into an auto–epistemic belief:[6]

$$\text{(AE–Translation 1)} \qquad (\mathbf{B}(A) \wedge \neg\mathbf{B}(B_1) \wedge \ldots \wedge \neg\mathbf{B}(B_N) \rightarrow C) \in K.$$

Notice that the consequence C of this auto–epistemic belief does not itself have to be a statement about the reasoner's beliefs. C may simply be a statement about the reasoner's environment (the world). Hence, we can view the belief $\mathbf{B}(A) \ldots \rightarrow C$ as part of the reasoner's theory about how his beliefs relate to the world, i.e., his theory (i) about his cognitive relationship to the world. Given AE–Transparency, C can also be replaced with $\mathbf{B}(C)$:

$$\text{(AE–Translation 2)} \qquad (\mathbf{B}(A) \wedge \neg\mathbf{B}(B_1) \wedge \ldots \wedge \neg\mathbf{B}(B_N) \rightarrow \mathbf{B}(C)) \in K.$$

The belief $\mathbf{B}(A) \ldots \rightarrow \mathbf{B}(C)$ is part of the reasoner's theory (ii) about how he draws inferences. AE–Transparency links both these theories (i) and (ii) closely together. Methodology thus becomes a matter of auto–epistemology.

The following observation summarizes our technical claims:

Observation 2.2. (D), AE–Translation 1 and AE–Translation 2 are mutually equivalent if AE–Transparency holds.

3. Diachronic Auto–Epistemology

The Diachronic Model

We will now proceed in the same three steps as in the synchronic case. First, we enrich the reasoner's language with auto–epistemic vocabulary that allows to describe the diachronic evolution of beliefs. Second, we discuss three diachronic coherence conditions on auto–epistemic beliefs. In view of Section 1, this step is the least secure. Finally, we discuss the diachronic coherence of methods for learning from evidence. I will not enter into the proof–theory of epistemic logics conforming to the discussed coherence conditions.

Let I be an interval of the integers serving as our diachronic index for both belief sets and belief operators. $\mathbf{B}_i(A)$ expresses the proposition that the reasoner believes A at stage i. This allows the reasoner to form beliefs about his own past, present, and future beliefs, e.g., $\mathbf{B}_j(A) \in K_i$ (with $i, j \in I$). Let $\mathcal{L}^{AE}$ be an auto–epistemically extended, propositional language that is closed under the classical truth–functional connectives as well as under the indexed operators $\mathbf{B}_i(.)$ (i.e., $\mathbf{B}_i(A) \in \mathcal{L}^{AE}$, for all $A \in \mathcal{L}^{AE}$ and all $i \in I$). Let $\mathcal{S}$ be the set of all subsets of $\mathcal{L}^{AE}$ that are consistent and closed with respect to classical propositional logic.

Any function $K_{(.)} : I \rightarrow \mathcal{S}$ is called an *epistemic trajectory*. K_i is the reasoner's belief set at i in the trajectory $K_{(.)}$. An *epistemic model* for the diachronic evolution of a reasoner's beliefs is a set $\mathcal{M} \subseteq \mathcal{S}^I$ of trajectories that, according to the model, the reasoner's beliefs might take. We may think of an epistemic model as, perhaps an observer's, cognitive theory about the reasoner, allowing the reasoner's actual trajectory to vary with the circumstances. Diachronic *coherence conditions* will apply to epistemic models by imposing constraints on which future trajectories of the model the reasoner himself can coherently deem possible from his present point of view. We consider epistemic models (i.e., sets of trajectories) because the coherence of present beliefs depends only on those future beliefs that are themselves coherent. We say that H_j is a (subjectively) *i–possible j–future* of $K_{(.)}$ in model $\mathcal{M}$ if and only if (i) $K_{(.)}, H_{(.)} \in \mathcal{M}$, (ii) H_j evolves from K_i (i.e., $K_i = H_i$), and (iii) H_j is subjectively possible from the viewpoint of K_i (i.e., $\neg\mathbf{B}_j(A) \notin K_i$ for all $A \in H_j$, and $\mathbf{B}_j(A) \notin K_i$ for all $A \notin H_j$). We shall throughout assume that a particular model $\mathcal{M}$ is given. For $K_{(.)} \in \mathcal{M}$, let $\mathcal{K}(i,j)$ be the set of all i–possible j–futures of $K_{(.)}$ in $\mathcal{M}$.

So far, we have not specified what leads a reasoner to change his beliefs. The current model can flexibly encompass any kind of belief change. To be more specific, however, let us also model how the reasoner learns from evidence. Let $\mathcal{EV}$ be a set of potential *evidence*. We assume that $\mathcal{EV}$ is a primitive parameter and follow AGM in stipulating $\mathcal{EV} \subseteq \mathcal{L}$ (where $\mathcal{L}$ may be $\mathcal{L}^0$ or $\mathcal{L}^{AE}$).[7] In concrete applications, evidence may represent the (propositional) output of the reasoner's sensory apparatus. Any function $E_{(.)} : I \rightarrow \mathcal{EV}$ is called an *evidential trajectory*. A *learning rule* is a function $\star : I \times \mathcal{S} \times \mathcal{EV} \rightarrow \mathcal{S}$ that, at any time, maps (prior) belief sets and (posterior) evidence into (posterior) belief set. $\star(i, K, A)$ is the belief set obtained from updating K with A at i. The AGM axioms are sometimes presented as a specification of desirable, or even logical, properties of learning rules (holding the index $i \in I$ fixed). Thus interpreted, the AGM theory of learning requires that learning rules satisfy the axioms (K*1)–(K*8). We will return to the interpretation of learning rules in Section 4. We shall throughout assume (K*6).

In a final addition to the reasoner's auto–epistemic vocabulary, we endow him with the capacity to reason about his past, present, and future evidence. Let $\mathbf{E}_i(A)$ express the proposition that the reasoner receives A as his (total) evidence at stage i ($A \in \mathcal{EV}$). We finally assume that $\mathcal{L}^{AE}$ is also closed under the operators $\mathbf{E}_i(.)$ (i.e., $\mathbf{E}_i(A) \in \mathcal{L}^{AE}$, for all $A \in \mathcal{EV}$ and all $i \in I$). We require that the reasoner's learning rule satisfy the demands of synchronic coherence, namely that the reasoner knows which evidence he has received:

(EV–Transparency)
$$\mathbf{E}_i(A) \in \star(i, K, A'), \text{ for all } A \Leftrightarrow A'.$$
$$\neg\mathbf{E}_i(A) \in \star(i, K, A'), \text{ for all } A \not\Leftrightarrow A'.$$

Let $\mathcal{EV}(K_{(.)}, i, j)$ be a set that contains, for any i–possible piece of $(i+1)$–evidence A (i.e., for any A with $\neg\mathbf{E}_{i+1}(A) \notin K_i$), exactly one member of its equivalence class $[A]$ (under classical propositional logic). When the reference to the trajectory $K_{(.)}$ is clear, we also write $\mathcal{EV}(i, j)$. We shall throughout assume that $\mathcal{EV}(K_{(.)}, i, j)$ is finite (for any $K_{(.)} \in \mathcal{M}$) and that the reasoner believes that he will receive (at least) one of the pieces of evidence that he currently deems possible.[8]

Belief changes in an epistemic model $\mathcal{M}$ are *evidence driven* if and only if for all trajectories $K_{(.)} \in \mathcal{M}$ there is some evidential trajectory $E_{(.)} \in \mathcal{EV}^I$ such that $K_{i+1} = \star(i+1, K_i, E_{i+1})$ (for all $i, i+1 \in I$). In evidence driven models, the prior subjective possibility of K_{i+1} depends on the prior subjective possibility of receiving E_{i+1} as evidence. Assuming that the reasoner knows at i which learning rule he uses at $i+1$, we obtain $\mathcal{K}(i, i+1) = \{\star(i+1, K_i, A) | \neg\mathbf{E}_{i+1}(A) \notin K_i\}$ (for $K_{(.)} \in \mathcal{M}$).

Generalized Reflection

The first auto–epistemic coherence condition that I wish to discuss is Generalized Reflection (for all $K_{(.)} \in \mathcal{M}$ and all $i \leq j$):[9]

(GR)
$$\bigcap \mathcal{K}(i, j) \subseteq K_i \subseteq \bigcup \mathcal{K}(i, j).$$

The left half of this principle requires that if A is believed in all subjectively possible futures, then A should already be believed today. The right half requires that if A is believed today, then there must be at least one subjectively possible future in which A continues to be believed.

I hope the reader will find this a plausible candidate for a diachronic coherence condition, for I have not much to offer in the way of conclusive argument (nor has the literature). In parallel to the synchronic case, this condition tries to substantiate the idea that an ideal belief state should not undermine itself. We may then say that a disbelief in A is undermined by the certainty of a future belief in A and that belief in A is undermined by the certainty of future disbelief in A.[10]

Let us examine the relationship between GR–coherence and AGM learning. Consider learning situations in which the reasoner is certain that he will receive either A or $\neg A$ as reliable evidence at $i+1$:

(Partitioning) $\quad (\mathbf{E}_{i+1}(A) \leftrightarrow \neg\mathbf{E}_{i+1}(\neg A)) \in K_i$

(Reliability) $\quad (\mathbf{E}_{i+1}(A) \rightarrow A), (\mathbf{E}_{i+1}(\neg A) \rightarrow \neg A) \in K_i$

Examples of such learning situations are laboratory experiments whose outcomes yield a reliable answer to a precise Yes–or–No question. Outside the laboratory, however, the reasoner regularly loses experimental control over his future evidence and, thus, the Partitioning assumption may fail.

At the heart of the AGM theory of learning are two axioms Inclusion/Expansion 1 ($K^\star3$) and Expansion 2 ($K^\star4$) that require evidence driven belief changes to be 'minimal'. ($K^\star3$) requires that if I learn a proposition A that is compatible with my prior beliefs (i.e., $\neg A \notin K_i$), then I add A and its consequences to my prior beliefs K_i. ($K^\star4$) requires that in this case I add nothing but A and its consequences. When Partitioning holds, learning rules that satisfy ($K^\star3$), ($K^\star4$) are GR–coherent. On the other hand, ($K^\star3$), ($K^\star4$) impose constraints on learning that go beyond GR–coherence. When Partitioning fails, however, ($K^\star3$) and ($K^\star4$) are insufficient to guarantee GR–coherence (and may even conflict with it; cf. Section 4).

Theorem 3.3. In evidence driven models:
(i) ($K^\star3$) plus Partitioning implies the left half of GR.
(ii) ($K^\star4$) plus Partitioning implies the right half of GR.
(iii) GR plus Partitioning implies neither ($K^\star3$) nor ($K^\star4$).

Binkley's Principle

Recall that the left half of GR requires that if A is believed in all subjectively possible futures, then A should be believed today. Binkley's (1968) Principle requires that the converse be also satisfied: A

should only be believed today if it is believed in all subjectively possible futures (for all $K_{(.)} \in \mathcal{M}$ and for all $i \leq j$):

$$\text{(BP)} \qquad K_i \subseteq \bigcap \mathcal{K}(i,j).$$

The combination of GR and BP yields $K_i = \bigcap \mathcal{K}(i,j)$ (what van Fraassen (1995) calls the 'Dogmatic Reflection Principle'). Binkley motivates his principle with the following example: Suppose you are about to leave your house in the morning. You are certain that in the evening you will believe that it will not rain on your way home. Imagine, for example, that your evening beliefs result from revising your morning beliefs with the new evidence received in the afternoon. Shouldn't you then leave your umbrella at home when you go out in the morning, i.e., shouldn't you already in the morning anticipate your evening beliefs and make it already part of your morning beliefs that it will not rain in the evening?

Binkley goes on to argue that BP "specifies the correct ideal if we are thinking of an ideal knower who uses his knowledge as a basis for planning for the future" (p.133). This defense obviously presupposes some connection between ungraded beliefs and action, but in the absence of an account of this connection, Binkley's defense remains at best a fragment (cf. Section 1). Again, I think that at present we can only resort to intuitions as preliminaries to a future theory.

Let us therefore consider an equivalent formulation of BP that perhaps more clearly expresses the idea that, in an ideal model $\mathcal{M}$, belief states must not be self–undermining (for all $K_{(.)} \in \mathcal{M}$ and all $i \leq j$):

$$\text{(Expected Monotonicity)} \qquad K_i \subseteq H, \qquad \text{for } H \in \mathcal{K}(i,j).$$

Don't assume your beliefs will turn out to be wrong!

Expected Monotonicity does not require that beliefs never be retracted. Rather, it requires the reasoner not to be satisfied with his current beliefs as long as they admit the possibility that they will be undermined and retracted in the future. In an ideal belief state, beliefs should be expected not to be retracted.

Theorem 3.4. In evidence driven models:
(i) ($K^\star 4$) plus Reliability implies BP.
(ii) GR, BP, Partitioning plus Reliability imply ($K^\star 1$)–($K^\star 6$), for all subjectively possible evidence.

Surprise Examination: GR and BP

Binkley discusses BP because it features in the Surprise Examination (SE) Paradox. Those who feel that SE is indeed paradoxical would reject either BP or GR. I agree with Binkley, however, that SE is not paradoxical. What is more, I argue that SE provides evidence for rather than against GR and BP. In the SE scenario, a teacher announces to her class that she will set an exam either on Monday or on Tuesday and that in either case the exam will take the class by surprise. If the students now obey both GR and BP, then they cannot coherently believe the teacher's announcement. It is this conclusion that has earned the scenario the name of a paradox.

Let us follow Binkley and first consider the synchronic version of the SE scenario. Suppose the teacher announces on Monday that she will set an exam on Monday and that this exam will take the class by surprise. In this case, the teacher's announcement is obviously self–undermining and violates synchronic coherence (in particular, Anti–Moore). Hence, the class cannot coherently believe the teacher's announcement.

In the full–blown diachronic version of SE we face a diachronically self–undermining announcement. It is the teacher's revelation that the exam will be set by Tuesday that undermines her claim that in no possible future will the class be able to anticipate the exam (Future 1: exam on Monday, Future 2: exam on Tuesday). BP–coherent reasoners must exclude the possibility that the surprise exam will be written on Tuesday (Future 2) because by Tuesday they would know that there has been no exam on Monday and and hence the Tuesday exam would not be a surprise anymore. (This is probably where the teacher went wrong. She probably did not meant to imply that the exam would still be a surprise at the last possible date.) GR–coherence next excludes the possibility that the surprise exam will be written on Monday (Future 1) because if by exclusion of Future 2 the only possible date for a surprise exam is Monday, then the Monday exam is not a surprise either.

Binkley in my view correctly concludes that the paradox "reduces to the phenomenon of incredible though possibly true propositions" (p. 136) and "belongs to the same family as Moore's [synchronic] paradox" (p. 135). I think that we can even use the SE scenario as a defense of GR and BP. Announcing the period into which an exam will fall and yet calling it a total surprise (even on the last day of that period) is diachronically X–incoherent, just as much as it is synchronically incoherent to announce the one precise date of a surprise exam. This, I think, should be a consequence of any notion X of diachronic coherence. Solving for X, we then find $X = \text{GR plus BP}$.[11] Thus, SE supports both GR and BP.

REFLECTION

Restricting our attention to evidence driven models, we consider the following principle that is equivalent to BP plus GR (for all $K_{(.)} \in \mathcal{M}$ and all $i, i+1 \in I$):

(Expansive Reflection) $K_i^+ \mathbf{E}_{i+1}(A) = \star(i+1, K_i, A)$,
for $\neg\mathbf{E}_{i+1}(A) \notin K_i$.

This formulation suggests a third (and final) coherence condition that draws on a notion of conditional belief:[12]

(Reflection) $K_i|\mathbf{E}_{i+1}(A) = \star(i+1, K_i, A)$.

If (K|4) holds, Reflection implies Expansive Reflection (i.e., GR and BP). Whereas GR and BP impose coherence conditions on unconditional beliefs, Reflection extends coherence to conditional auto–epistemic beliefs: Under the supposition that you are to believe A and disbelieve B in the future, you should believe A and disbelieve B. The idea behind this condition is again that beliefs — now including conditional beliefs — should ideally not undermine each other. Under the supposition that you form a future belief or disbelief, you should ideally see no reason to doubt that this belief or disbelief is justified. You should now fully commit yourself to the way in which you change your beliefs in the future.

Theorem 3.5. In evidence driven models, GR plus BP is equivalent to Expansive Reflection.

Theorem 3.6. Assume Reflection, (K|3), (K|4), (K|7), and (K|8).
Then (*) $A \in K_i|\mathbf{E}_{i+1}(A)$ plus (**) $\mathbf{E}_{i+1}(A) \in K_i|A$ implies (K*1)–(K*6).

4. Learning from Evidence

AUTO–EPISTEMOLOGY AND LEARNING

Let me explain a view on learning from evidence that I call actualism. In hindsight, it may be easy to scold this view for its naivety from which it seems to derive considerable appeal and popularity. According to actualism, learning from evidence is insensitive to the epistemic processes that connect the reasoner to the world. In actualist learning, it suffices simply to take the received piece of evidence $A \in \mathcal{EV}$ and to feed it mechanically into a learning rule $\star$ (a function of prior beliefs and the actual piece of evidence).[13] There is no need for the reasoner to form beliefs about his cognitive relationship to the world. In particular, the reasoner needs no beliefs about what process (e.g., a measurement) generates the evidence, i.e., no beliefs about what evidence different from the actual he might have received under different circumstances (hence, my term 'actualism'). The simplicity of this view is striking.[14]

A reasoner with auto–epistemic capacities is more sophisticated. He forms subjective assessments of his evidence and his learning situation. The Partitioning condition in Section 3 was an example of such an assessment. In Section 2, we have already come across the importance of auto–epistemic beliefs for default reasoning. Here we consider their importance for learning from evidence. Beliefs like the Partitioning assessment express (i) the reasoner's personal theory about his epistemic relationship to the world and about the processes that generate his evidence. They express (ii) his personal theory about how he learns from evidence. Diachronic coherence binds these two theories (i) and (ii) together and methodology becomes once more a matter of auto–epistemology.

Coherence conditions thus furnish a manual for translating axiomatic properties of learning rules into different auto–epistemic assessments of a given learning situation. I believe it is a virtue of this auto–epistemic model that it can make sense of the multiplicity of learning rules that have been suggested in the literature (AGM, Katsuno/Mendelzon's (1992) Updating, Fuhrmann's (1997) Merge etc.). Coherence conditions afford a logic for reasoning about these learning rules.

Coherent Learning

A sophisticated AE–reasoner can no longer enjoy the simple pleasures of actualism. Under Reflection, different auto–epistemic assessments prescribe the use of different (actualist) learning rules. The reasoner is coherent in using a particular learning rule if and only if he makes the corresponding prior assumptions about his posterior evidence. On the one hand, GR– and BP–coherence therefore justifies the basic AGM axioms ($K^\star$1)–($K^\star$6) when the reasoner makes the Partitioning and the Reliability assumption (Theorem 3.4). On the other hand, I can see no reason why such prior assumptions should be universally the same in all possible learning situations. I therefore do not believe that the use of one universal, actualist learning rule can be coherent.

Take the example of the Success (K*2) condition ($A \in K_i^\star A$). It is coherent if and only if the reasoner actively believes in the Reliability of his evidence. Mere suspense of this belief — short of active belief in the deceptiveness of evidence — already suffices to render Success incoherent. (This illustrates what extreme caution BP–coherence requires when accepting new information.) At the Konstanz conference Sven–Ove Hansson provided us with an à la mode example: When we learn from a child that some vase was broken by a dinosaur, what do we infer?

Let us return to (K*3), (K*4) as a characterization of 'minimal' belief changes. (K*4) bars new[15] inductive inferences from A that would go beyond what A and K_i jointly imply. (K*3) excludes direct disconfirmation of a prior belief. Prior beliefs can only be disconfirmed indirectly through the confirmation of logically conflicting hypotheses.[16] GR and BP generally allow belief changes that are not minimal. Non–minimal belief changes and inductive inferences try to extrapolate excess information from a given piece of evidence A. Such extrapolations are coherent precisely when the reasoner believes that there is something additional to be learnt from A apart from its consequences. When Partitioning and Reliability hold, we have a case where the reasoner believes that there is nothing to be learnt apart from A. For then we have $(A \leftrightarrow \mathbf{E}_{i+1}(A)) \in K_i$ and, by Theorem 3.4, (K*3) and (K*4) are coherent.

A famous example of an actualist learning rule is the identification of the beliefs after learning A with the beliefs conditional on A:[17]

(Conditionalization) $\quad \star(i+1, K_i, A) = K_i|A$

Is Conditionalization a coherent, actualist learning rule? The answer to this question depends on whether or not $K_i|A = K_i|\mathbf{E}_{i+1}(A)$. Again, there is no reason why this equation should hold universally. It puts severe restrictions on prior conditional beliefs. I therefore think it is wrong to assume that coherent learning generally goes by Conditionalization.

It is sometimes suggested that we modify our notion of evidence. Instead of using A as the input for a learning rule $\star$, it is suggested that we rather use the proposition "I learn A at j"/"My evidence at j is A". Posterior beliefs would then be given by $K_i^\star \mathbf{E}_{i+1}(A)$. This suggestion implicitly acknowledges that before the reasoner can learn from A, he needs a sophisticated theory about the process that has generated A as its output (e.g., a measurement). Hence, this suggestion does not view Conditionalization as an actualist learning rule, but in effect reinterprets it as the auto–epistemic Reflection principle.[18]

Summary

I have tried to explore some reasons for accepting diachronic coherence principles for auto–epistemic reasoning. On the basis of such principles, properties of acceptable learning rules translate into (conditional) prior beliefs. If consistency is the only constraint on prior beliefs, then consistency is also the only constraint on acceptable learning rules. But then consistency is also the only constraint on posterior beliefs. Conversely, if the axioms of AGM, Updating, or Merge are accepted as constraints on learning rules, then there must be constraints on prior beliefs over and above consistency.

The multiplicity of learning rules in the literature suggests that there is no agreement on what these additional constraints could be. I do not believe that AGM or any other axiom system can specify a universally applicable learning rule (except for the Logical Core), although they may be justified in special cases (Partitioning, Reliability). I have also tried to argue that the AGM axioms are plausible as axioms for conditional belief, but that conditional belief and learning rules do not generally coincide. On the view defended here, a reasoner must have a prior assessment of his evidence before he can even learn from this evidence. Hence, a position is untenable that uses a universal theory of learning in order to counter-act the subjectivism that results from weak, coherentist, constraints on prior beliefs.

Appendix

Theorem 3.3: Assume that $K_{(.)} \in \mathcal{M}$. Holding $i \in I$ fixed, we write $K_i^\star A := \star(i+1, K_i, A)$. Partitioning implies $\mathcal{K}(i, i{+}1) = \{K_i^\star A, K_i^\star \neg A\}$. (i) Suppose $B \in K_i^\star A$ and $B \in K_i^\star \neg A$. By (K*3), $K_i^\star A \subseteq K_i^+ A$ and $K_i^\star \neg A \subseteq K_i^+ \neg A$. Hence, $(A \to B) \in K_i$ and $(\neg A \to B) \in K_i$. By the closure of K_i, we obtain $\neg B \in K_i$. (ii) Suppose $B \notin K_i^\star A$ and $B \notin K_i^\star \neg A$. By the consistency of K_i, either $A \notin K_i$ or $\neg A \notin K_i$. Suppose $A \notin K_i$. By (K*4), $K_i^+ \neg A \subseteq K_i^\star \neg A$. Since $B \notin K_i^\star \neg A$, $(\neg A \to B) \notin K_i$. By the closure of K_i, we obtain $B \notin K_i$. Similarly for $\neg A \notin K_i$. (iii) Construct counter–example.

Theorem 3.4: (i) Suppose that for some $H \in \mathcal{K}(i, i+1)$, $B \notin H$. H must be of the form $K_i^\star A$ with $\neg \mathbf{E}(A) \notin K_i$. Hence, by the closure of K_i and by Reliability, $\neg A \notin K_i$. Since $B \notin K_i^\star A$ and since by (K*4), $K_i^+ A \subseteq K_i^\star A$, we have $(A \to B) \notin K_i$. By the closure of K_i, we obtain $B \notin K_i$. (ii) Use Expansive Reflection (Theorem 3.5) to translate (K*1)–(K*6) into prior beliefs and check that they follow from Partitioning plus Reliability.

Theorem 3.5: In evidence driven models, GR plus BP takes the form $K_i = \bigcap_{A \in \mathcal{EV}(i,i+1)} \star(i+1, K_i, A)$. a) We show that GR plus BP implies that $(\mathbf{E}_{i+1}(A) \to B) \in K_i$ iff $B \in \star(i+1, K_i, A)$ for $\neg \mathbf{E}_{i+1}(A) \notin K_i$ (Expansive Reflection). a1) Assume $\neg \mathbf{E}_{i+1}(A) \notin K_i$ and $(\mathbf{E}_{i+1}(A) \to B) \in K_i$. By BP, $(\mathbf{E}_{i+1}(A) \to B) \in \star(i+1, K_i, A)$. By EV–Transparency, $\mathbf{E}_{i+1}(A) \in \star(i+1, K_i, A)$ and thus $B \in \star(i+1, K_i, A)$. a2) Assume $B \in \star(i+1, K_i, A)$. Then $(\mathbf{E}_{i+1}(A) \to B) \in K_i$ (by GR) because for all A', $(\mathbf{E}_{i+1}(A) \to B) \in \star(i+1, K_i, A')$. [If $A \not\Leftrightarrow A'$, by EV–Transparency $\neg\mathbf{E}_{i+1}(A) \in \star(i+1, K_i, A')$ and thus $(\mathbf{E}_{i+1}(A) \rightarrow B) \in \star(i+1, K_i, A')$. If however $A \Leftrightarrow A'$, then by assumption $B \in \star(i+1, K_i, A')$ and thus again $(\mathbf{E}_{i+1}(A) \rightarrow B) \in \star(i+1, K_i, A')$.] b) Expansive Reflection yields GR and BP because (*) $K_i = \bigcap_{A \in \mathcal{EV}(i,i+1)} K_i^+\mathbf{E}_{i+1}(A)$. To prove (*), notice that $B \in \bigcap_{A \in \mathcal{EV}(i,i+1)} K_i^+\mathbf{E}_{i+1}(A)$ iff (**) $(\mathbf{E}_{i+1}(A) \rightarrow B) \in K_i$ for all $A \in \mathcal{EV}(i, i+1)$. Since we assume $\mathcal{EV}(i, i+1)$ to be finite, $\bigwedge_{A \in \mathcal{EV}(i,i+1)}(\mathbf{E}_{i+1}(A) \rightarrow B) \in K_i$. From our assumption $\bigvee_{A \in \mathcal{EV}(i,i+1)} \mathbf{E}_{i+1}(A) \in K_i$, we obtain $B \in K_i$ iff (**).

Theorem 3.6: It is well–known that in the presence of (K|7) and (K|8), (*) plus (**) implies $K_i|A = K_i|\mathbf{E}_{i+1}(A)$, whence by Reflection $K_i^\star A = K_i|A$. Thus, ($K^\star$2) follows from (*) and ($K^\star$3), ($K^\star$4) follow from (K|3), (K|4), respectively.

Notes

[1] The Bayesian literature introduces a notion of 'full belief' (propositions with maximal subjective probability) but argues that epistemic states are more completely characterized by some subjective probability measures. In the AGM literature, 'epistemic entrenchment' is another possible aspect of epistemic states, underlying the dynamic evolution of (ungraded) belief sets.

[2] I use | as a symbol for conditional belief and reserve the symbol $\star$ for rules for learning from evidence. I write $(K|1)$ etc. when I talk about axioms for conditional belief. We write $A \Leftrightarrow B$ iff A and B are equivalent with respect to classical propositional logic. We write $\top$ for any sentence of $\mathcal{L}$ that is tautological and $\bot$ for any sentence of $\mathcal{L}$ that is contradictory (with respect to classical propositional logic). K^+A is the closure of $K \cup \{A\}$ under classical propositional logic.

[3] I am here talking about how the AGM axioms are typically motivated, notwithstanding AGM's prominent results on conditional beliefs and beliefs in conditionals.

[4] I have baptized the central axiom 'Multiplication' because it corresponds under the $p = 1$ definition to the Multiplication axiom for conditional probabilities. This axiom is part of a characterization of probabilities conditional on events with zero probability. (Such conditional probabilities cannot be reduce to the usual ratio of unconditional probabilities.)

[5] Proof: If $\neg A \notin K|C$, then $K|C|A = (K|C)^+A$ by $(K|3)$, $(K|4)$. Thus, (N) follows from $(K|7)$, $(K|8)$. Now use Theorem 1.1.

[6] Konolige (1987) applies AE–Transparency in a proof–theoretic analysis of default reasoning.

[7] AGM does not explicitly discuss the (auto–epistemic) richness of $\mathcal{L}$. On another occasion, we shall discuss the notion of coherence when evidence (as well as epistemic states) are endowed with a structure richer than propositions. Cf. Hild (1998a) for coherence in the Bayesian model.

[8] I.e., $\bigvee_{A \in \mathcal{EV}(i,j)} \mathbf{E}_j(A) \in K_i$ (given finite $\mathcal{EV}(i,j)$) which follows from GR (cf. below) plus EV–Transparency.

[9] Under the $p = 1$ definition (Section 1), GR is a special case of van Fraassen's (1995) probabilistic General Reflection Principle.

[10] An obvious qualification is that future beliefs are assumed to be endorsable or rational from the current point of view.

[11] Binkley uses GR implicitly. It is the left half of GR that implies: If $\mathbf{B}_j(A \rightarrow B) \in K_i$, then $(\mathbf{B}_j(A) \rightarrow \mathbf{B}_j(B)) \in K_i$.

[12] Under the $p = 1$ definition, Expansive Reflection is a special case of van Fraassen's (1984) probabilistic Reflection principle. What I call Reflection in this paper goes beyond this principle and cannot be justified by a decision–theoretic coherence argument even if we accept the $p = 1$ definition.

[13] Levi (1980) prefers to speak of 'input' instead of 'evidence' in this context. 'Evidence', for Levi, is the proposition that A was received as input. I think this is the exact reverse of my usage.

[14] Its simplicity is compromised perhaps only by the assumption that the epistemic process must be believed to generate reliable, true evidence — an assumption that is fit to exclude simple–minded scepticism.

[15] Prior inductive hypotheses may already be contained in K_i.

[16] This is also a consequence of Fuhrmann's (1997) weaker Relevance axiom.

[17] If we assume that conditional beliefs satisfy (K|1)–(K|8), then Conditionalization implies (K*1)–(K*8).

[18] The suggested re-interpretation of Conditionalization reads '$K_i^\star \mathbf{E}_{i+1}(A) = K_i|\mathbf{E}_{i+1}(A)$'. Since $K_i^\star \mathbf{E}_{i+1}(A)$ is suggested to represent the reasoner's posterior beliefs after having learnt A, this equation is identical to Reflection.

References

Binkley, R. (1968), "The Surprise Examination in Modal Logic", *J Phil* **65**, 127-136

Fuhrmann, A. (1997), *An Essay on Contraction*, Stanford: CSLI Publications

Gärdenfors, P. (1988), *Knowledge in Flux*, Cambridge (Mass.)/London: MIT

Hild, M. (1998a), "The Coherence Argument Against Conditionalization", *Synthese* **115**, 229–258.

Hild, M. (1998b), "Auto–Epistemology and Updating", forthcoming in *Philosophical Studies.*

Hintikka (1962), *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, Ithaca (NY): Cornell University

Katsuno, H., Mendelzon, A.O. (1992), "On the Difference between Updating a Knowledge Base and Revising it", in: P. Gärdenfors (ed.), *Belief Revision*, Cambridge: CUP, 183–203

Konolige, K. (1987), "On the relation between Default and Autoepistemic Logic", in: M.L. Ginsberg (ed.), *Readings in Nonmonotonic Logic*, Los Altos (CA): Morgan Kaufmann, 195-226

Levi, I. (1980), *The Enterprise of Knowledge*, Cambridge (Mass.): MIT

Reiter, R. (1980), "A Logic for Default Reasoning", *Artificial Intelligence* **13**, 81–132

Rescher, N. (1964), *Hypothetical Reasoning*, Amsterdam: North–Holland

Stalnaker, R.C. (1980), "A Note on Nonmonotonic Modal Logic", Department of Philosophy, Cornell University

van Fraassen, B.C. (1984), "Belief and the Will", *J Phil* **81**, 235-256

van Fraassen, B.C. (1995), "Belief and the Problem of Ulysses and the Sirens", *Philosophical Studies* **77**, 7-37

Address for correspondence: Christ's College, Cambridge CB2 3BU (UK), email: mh247@cam.ac.uk