

# Auto–Epistemology and Updating

Matthias Hild

*Balliol College, Oxford*

17 February, 1997

**Abstract.** The first and second section of this paper extend auto–epistemology from the context of full belief into a probabilistic framework with Reflection as its central principle. The third and fourth section examine how auto–epistemology relates to the choice of update methods, especially Conditionalization. This update rule has to be distinguished from the superficially similar, but logically independent, principle of Auto–Epistemic Conditionalization. Possible violations of Conditionalization are discussed. The final section develops the auto–epistemology of full belief in structural correspondence to the probabilistic case. The diachronic principle corresponding to Reflection appears in the Surprise Examination Paradox. AGM updates correspond to Conditionalization and are auto–epistemically validated only when the reasoner’s subjective assessments of her learning situation meets particular conditions.

**Key words:** AGM, Auto–Epistemology, Belief Revision, Conditionalization, Freund’s Puzzle, Reflection, Surprise Examination Paradox, Updating

## Introduction

Consider the example of a robot that is equipped with a sensory apparatus. In describing such a robot *qua* cognitive system, we ascribe to it epistemic states that, on the one hand, are sensitive to changes in the robot’s environment and, on the other, influence its practical reactions to such changes. Our epistemological theory about the robot will attempt to develop a framework of epistemic states and evidential input in which we can model the system’s cognitive behaviour. Within such a framework, we may, for example, examine the reliability of the robot’s evidential input and to what extent the robot forms correct or incorrect representations of its environment. Investigations into the robot’s methodology will, finally, examine how it draws conclusions and how it updates its epistemic states in reaction to evidential input.

In the same way in which an observer can form an epistemological theory about another system, like this robot, auto–epistemic reasoners can form such theories about themselves. Auto–epistemic opinions of a reasoning system are opinions about its own epistemic condition. Typical examples are opinions about one’s own opinions. In particular, auto–epistemology enables a system to reason about its own methodology. In non–probabilistic models, this connection has so far only been explored in application to non–monotonic default reasoning.<sup>1</sup> It is the merit of the literature on probabilistic updating that it has brought

attention to the connection between diachronic auto-epistemology and update methods. Unfortunately, the overall conception of probabilistic auto-epistemology remained fragmentary and has led to numerous misunderstandings and logical errors. Section 1 and Section 2 of this paper will therefore try to supply a rigorous development of probabilistic auto-epistemology centred around van Fraassen's (1983), (1984) Reflection principle and Goldstein's (1983) Iteration. These diachronic principles require that present probabilities equal the expectation (weighted average) of future probabilities.

It is important to realize that these auto-epistemic principles are not committed to any particular update method, let alone to Conditionalization. Differences in the auto-epistemic assessment of future evidence will cause Reflection to prescribe different methodological reactions to this evidence. Auto-epistemic reasoning may, in particular, enforce violations of Conditionalization. If understood as a universal update methodology, Conditionalization is therefore at odds with auto-epistemic reasoning. In Section 3, I will clarify this situation and address widespread misconceptions about the logical connections between Reflection and Conditionalization (e.g., Maher (1993) and van Fraassen (1995)).<sup>2</sup> I will argue that these misconceptions stem from a confusion of Conditionalization with the non-methodological principle of 'Auto-Epistemic Conditionalization'. An analysis of Freund's (1965) Puzzle in Section 4 will further illustrate this point.

In the context of full belief, auto-epistemology has mostly been treated as a modal logic with synchronic iteration principles for the belief operator as central axioms (cf. Hintikka (1962)). This focus on the syntactical aspects of auto-epistemology might explain why the connection between auto-epistemology and updating has not yet been investigated in models of full belief. In order to explore this connection, I will therefore in Section 5 develop an auto-epistemic model of full belief in structural correspondence to its probabilistic counterpart. (Full beliefs structurally correspond to maximal probabilities.) It turns out that in models of full belief, Reflection corresponds to a principle that already appears in the Surprise Examination Paradox. I therefore discuss a non-paradoxical interpretation of surprise examinations.

The probably most prominent model for updating full beliefs has been developed by Alchourrón, Gärdenfors, and Makinson (AGM model). AGM updates aspire to minimizing the changes in prior beliefs that are necessary in order to incorporate new information consistently. It will become clear that AGM updating structurally corresponds to Conditionalization. As was the case with Conditionalization, auto-epistemic reasoning undercuts the universalistic claims of the AGM update methodology. The AGM axioms must therefore be understood

as a characterization of conditional belief (arguably, identical to the operation of supposing) rather than of an update methodology.

The investigations in this paper aim to develop the following insight: In auto-epistemic reasoning, the subjective assessment of a given learning situation (opinions about the nature and quality of evidence) is systematically prior to the choice of a suitable update method. Auto-epistemic reasoning about particular learning situations is therefore at odds with universal update methodologies such as Conditionalization or AGM updating.

## 1. Probabilistic Model

### PROBABILITIES

In the present framework, opinions (or, probabilistic information) are represented by unique probability functions. Let  $\mathcal{A}$  be a  $\sigma$ -algebra over a set of possible worlds  $\Omega$ . Let  $P_i(\cdot)$  be a probability measure on  $\mathcal{A}$ . I shall also refer to  $\mathcal{A}$  as the reasoner's 'language' since  $\mathcal{A}$  represents the reasoner's ability to discriminate between different possible worlds. Let the diachronic index  $I$  be a discrete and totally ordered set that indexes different stages of reasoning.<sup>3</sup> The sequence  $\langle P_i \rangle_{i \in I}$  then describes the diachronic changes in probabilities over different stages of reasoning. The totality of such sequences describes the possible trajectories along which the reasoner's opinion can evolve. In order to simplify the following presentation, I will restrict the discussion to probability measures that assign a non-zero probability only to countably many worlds from  $\Omega$ . We can thus define the expectation  $E_{P(\cdot)}(X) := \sum_{\omega \in \Omega} P(\omega)X(\omega)$  of a random variable  $X : \Omega \rightarrow \mathfrak{R}$  without using integration theory. The proofs in this paper can be quite easily extended to the more general setting.

Conditional probabilities  $P_i(\cdot|\cdot)$  are characterized by the following axioms:<sup>4</sup>

- (CP1)  $P_i(\cdot|A)$  is a probability measure, if  $A \neq \emptyset$ .  
(Normality 1)
- (CP2)  $P_i(A|A) = 1$ , if  $A \neq \emptyset$ .  
(Normality 2)
- (CP3)  $P_i(A \cap B|C) = P_i(A|B \cap C) \times P_i(B|C)$   
(Multiplication)
- (CP4)  $P_i(\cdot|\Omega) = P_i(\cdot)$   
(Reduction)

$\mathcal{PR}_{\mathcal{A}}$  is the set of all conditional probability measures over  $\mathcal{A}$  satisfying axioms (CP1)–(CP4). In the presence of Reduction, the Multiplication axiom implies that conditional probabilities blend into unconditional probabilities in the following manner:

$$(CP5) \quad P_i(A \cap B) = P_i(A|B) \times P_i(B)$$

Hence,

$$P_i(A|B) = \frac{P_i(A \cap B)}{P_i(B)}, \quad \text{if } P_i(B) > 0. \quad (1)$$

Probabilities conditional on events with non-zero probability therefore reduce to the ratio of unconditional probabilities.

#### CONDITIONALIZATION

Incoming evidence will rarely determine a new probability function uniquely. Standard probabilistic methodology therefore uses certain features of the old probability function to supplement the new constraint. These are the features that are thought to remain undisturbed by the incoming evidence. We start with an ‘original’  $\sigma$ -algebra  $\mathcal{A}^0$  over  $\Omega$  that we will in the next section enrich by auto-epistemic vocabulary. Assume that the evidential input specifies new probabilities  $p_k$  for the elements of a partition  $\{B_k | k \in K\} \subseteq \mathcal{A}^0$  of  $\Omega$ . We call  $\langle B_k, p_k \rangle_{k \in K}$  a *Jeffrey constraint* over  $\mathcal{A}^0$ . If we can furthermore assume that the conditional probabilities  $P_i(\cdot|B_k)$  remain constant during the update with the new values for the  $B_k$ , then the new probability measure is uniquely determined:

$$(Gen.Cond) \quad [P_i^* \langle B_k, p_k \rangle_{k \in K}](A) = \sum_{k \in K} p_k \times P_i(A|B_k),$$

for countable  $K$  and  $A \in \mathcal{A}^0$ .

This is Jeffrey’s (1983) understanding of the scope of Generalized Conditionalization. It holds only under (and is in fact equivalent to) the following two conditions:

$$(Success) \quad [P_i^* \langle B_k, p_k \rangle_{k \in K}](B_{k'}) = p_{k'}, \quad k' \in K.$$

$$(Rigidity) \quad [P_i^* \langle B_k, p_k \rangle_{k \in K}](A|B_{k'}) = P_i(A|B_{k'}),$$

where  $k' \in K$  and  $A \in \mathcal{A}^0$ .

We obtain Conditionalization as the special case in which some  $B \in \mathcal{A}^0$  is assigned maximal, and  $\neg B$  minimal, probability:

$$(Conditionalization) \quad [P_i^* B](A) = P_i(A|B), \quad \text{for } A \in \mathcal{A}^0.$$

It is important to distinguish sharply between Conditionalization and the (partial) definition (1) of conditional probabilities. Hacking (1967)

had to argue this obvious point. In the above formulation of (Generalized) Conditionalization, conditional probabilities merely function as a shorthand for the above ratio of unconditional probabilities (as long as  $P_i(B_k) > 0$ ). Conditionalization, on the other hand, goes beyond a mere abbreviation for the sake of notational convenience. It imposes non-trivial, substantive, restrictions on the relationship between prior and posterior probability measures (formulated in terms of conditional probabilities).

Under Jeffrey's (1983) interpretation, (Generalized) Conditionalization is less of an update rule than a probabilistic tautology: If your updated probabilities are to fulfil Success and Rigidity, then they are determined by Generalized Conditionalization. Understood as a genuine update rule, however, (Generalized) Conditionalization amounts to the two *prescriptions* that updated probabilities should always satisfy Success and Rigidity. That Success should always hold is more or less taken for granted whereas a large number of arguments have been offered in support of Rigidity (most famously, the so-called Dutch book arguments, cf. Teller (1973)).

## 2. Probabilistic Auto-Epistemology

### DIACHRONIC COHERENCE

Auto-epistemic vocabulary allows a reasoner to form an opinion about her own epistemic condition because it enables her to construct her own epistemological theory about herself. She can, in particular, reason about how to choose an update mechanism for her current probabilities if she faces a certain learning situation. The principles of diachronic auto-epistemology prescribe that she should choose an update mechanism such that the expected value of her future probabilities under this update mechanism equal her current probabilities. This prescription can be defended by a Coherence Argument (Dutch Book Argument), notwithstanding the failure of such arguments for the classical update rules (cf. Hild (1997)). Roughly, an expected-utility maximizer will incur a sure loss if and only if she violates Reflection. This result indicates a diachronic imbalance in probabilities that violate Reflection.

Van Fraassen (1984) consequently argues that a violation of Reflection would 'undermine [one's] own status as a person of [cognitive] integrity'. He illustrates this view by the example of a weather forecaster. The forecaster, firstly, announces on Monday morning (personal) probabilities for rain on Tuesday. Secondly, she specifies probabilities for the weather conditions that could occur during Monday. Thirdly,

she writes up a manual of how she would on Tuesday morning adjust the probabilities for rain depending on the observed weather condition on Monday. Reflection applies to how these three components of the forecaster's opinions should relate to each other. It recommends the choice of an update manual such that on Monday morning, the forecaster's probability for rain on Tuesday is identical to the expectation of her updated probability on Tuesday morning (where this updated probability is determined on the basis of the observed weather conditions during Monday and the methods laid out in the manual).

In technical terms, the forecaster has not merely constructed a static Kolmogorov–model for rain on Tuesday. Her updating manual rather constructs a whole tree of Kolmogorov–models along a time–axis (Monday, Tuesday).<sup>5</sup> The first of these models describes her probabilities on Monday. The subsequent branches result from updating Monday's probabilities. Each of these branches is assigned a certain probability on Monday (using Monday's probabilities for the weather on Monday and the updating manual). If we calculate Monday's expectation of the probability that will be assigned to  $A$  at the Tuesday nodes, we should, according to Reflection, arrive at Monday's probability for  $A$ .

Reflection is the probabilistic version of an idea that has first been proposed in the context of full belief: We should already fully believe today what we are convinced we will fully believe tomorrow. What is more, it also generalizes the converse principle suggested by Binkley (1968) in connection with the Surprise Examination Paradox: Ideally, we should only fully believe today what we are convinced we will fully believe tomorrow. Binkley argues that, although only ideally rational agents could always meet these demands, they are necessary for successfully planning the future. Suppose that I am about to leave my house in the morning. I am also certain that in the evening I will believe that it will not rain. If I thought that I would be irrational in the evening, tonight's beliefs should not matter for my beliefs in the morning. If I however think that in the evening I will be justified to believe in dry weather, Binkley's Principle implies that I should already in the morning believe that it will not rain in the evening. Otherwise, I might in the evening end up carrying an umbrella with the firm belief that it will not rain.<sup>6</sup>

#### REFLECTION, TRANSPARENCY, AND PERFECT MEMORY

Let me now introduce the elements of a formal model of an auto–epistemic reasoner. As before, let  $\Omega$  be a non–empty set of possible worlds, let  $\mathcal{A}$  be a  $\sigma$ –algebra over  $\Omega$ , and  $I$  a diachronic index. An *epistemic function* relative to  $\mathcal{A}$  is a function  $P_{(\cdot,\cdot)} : \Omega \times I \rightarrow \mathcal{PR}_{\mathcal{A}}$

that attaches a probability measure over  $\mathcal{A}$  to each possible world  $\omega$  at each stage  $i$ . Epistemic functions describe the reasoner's subjective probabilities at a certain stage  $i$  under the external conditions  $\omega$ . The reasoner can hence give her probability for the event that she will have a certain personal probability in the future. Let  $\text{SP}_i(Q) := \{\omega | P_{\omega,i}(\cdot) = Q(\cdot)\}$  express that the reasoner's subjective probability function at  $i$  equals  $Q$  (where  $Q \in \mathcal{PR}$ ). We start from the original  $\sigma$ -algebra  $\mathcal{A}^0$  over  $\Omega$  that we will now enrich by auto-epistemic vocabulary. We say that  $\langle \Omega, \mathcal{A}^{AE}, I, P_{(\cdot,\cdot)} \rangle$  is an *auto-epistemic model* if and only if  $P_{(\cdot,\cdot)}$  is an epistemic function relative to  $\mathcal{A}^{AE}$  and  $\mathcal{A}^{AE}$  is the closure of  $\mathcal{A}^0$  under the auto-epistemic vocabulary  $\text{SP}_i(Q)$  (for all  $i \in I$  and  $Q \in \mathcal{PR}_{\mathcal{A}^{AE}}$ ).

Diachronic auto-epistemic principles will impose restrictions on what epistemic functions  $P_{(\cdot,\cdot)}$  are admissible. Let us begin with the central principle of Reflection. Its proper formulation requires the notion of the possible posterior probability measures into which a given prior probability measure may evolve, or branch. Let us call  $P_{(\omega,\cdot)}$  the *epistemic trajectory* in  $\omega$ . An epistemic trajectory specifies a sequence  $\langle P_{(\omega,\cdot)} \rangle_{i \in I}$  of probability measures across the diachronic index  $I$ . We say that the trajectory  $P_{(\omega_2,\cdot)}$  is a *possible  $i$ -branch* of  $P_{(\omega_1,\cdot)}$  if and only if  $P_{\omega_1,i} = P_{\omega_2,i}$ . If  $P_{(\omega_2,\cdot)}$  is a possible  $i$ -branch of  $P_{(\omega_1,\cdot)}$  and  $i \leq j$ , then  $\text{SP}_j(P_{\omega_2,j}) = \{\omega | P_{\omega,j} = P_{\omega_2,j}\}$  is the event that the epistemic trajectory  $P_{\omega_1,\cdot}$  branches into a trajectory with the probability measure  $P_{\omega_2,j}$  at  $j$ .

Reflection requires that the probability in the  $\omega_1$ -trajectory given a future branching into  $P_{\omega_2,j}$  equals the future probability  $P_{\omega_2,j}$ :<sup>7</sup>

$$\begin{aligned} \text{(Reflection)} \quad & \forall i, j \ (i \leq j) \quad \forall \omega_1, \omega_2 \ (P_{\omega_1,i} = P_{\omega_2,i}) : \\ & P_{\omega_1,i}(\cdot | \text{SP}_j(P_{\omega_2,j})) = P_{\omega_2,j}(\cdot). \end{aligned}$$

To simplify this notation, I will from now on only refer to two trajectories  $P_i := P_{\omega_1,i}$  and  $Q_i := P_{\omega_2,i}$  ( $i \in I$ ). Statements involving these two trajectories therefore implicitly contain a universal quantification over worlds  $\omega_1$  and  $\omega_2$ . Reflection then reads:

$$\begin{aligned} \text{(Reflection)} \quad & \forall i, j \ (i \leq j) \quad P_i = Q_i : \\ & P_i(\cdot | \text{SP}_j(Q_j)) = Q_j(\cdot). \end{aligned}$$

By 'Non-Zero Reflection' ('NZ-Reflection') I mean the restriction of Reflection to cases where  $P_i(\text{SP}_j(Q_j)) > 0$ .

In order to keep auto-epistemic reasoning tractable, let us exclude any uncertainties and errors on the reasoner's part about her present opinion:

$$\text{(AE-Transparency)} \quad P_i(\text{SP}_i(Q_i)) = \begin{cases} 1, & \text{if } P_i(\cdot) = Q_i(\cdot). \\ 0, & \text{otherwise.} \end{cases}$$

AE-Transparency postulates a Cartesian perfection of introspection that presupposes a high degree of idealization. Where AE-Transparency fails, auto-epistemic reasoning will also fail simply because auto-epistemic reasoning theorizes about the epistemological condition of a cognitive system from the point of view of that system. AE-Transparency therefore expresses a preoccupation with situations where auto-epistemic reasoning actually succeeds.

AE-Transparency can be extended into Perfect Memory of the past:

$$\text{(Perfect Memory)} \quad \forall i, k \ (k \leq i) : \\ P_i(\text{SP}_k(Q_k)) = \begin{cases} 1, & \text{if } P_k(\cdot) = Q_k(\cdot) \\ 0, & \text{otherwise} \end{cases}$$

*Theorem 2.1.* (i) Reflection implies AE-Transparency.

(ii) Reflection implies Perfect Memory.

Proofs will always be given in the Appendix.

#### MORE AUTO-EPISTEMIC PRINCIPLES

Reflection is not the only diachronic principle that has been proposed for auto-epistemic probabilities. Chronologically, it was preceded by Spohn's (1978) and Goldstein's (1983) requirement that present probabilities (or, expectations) should be identical to the present expectation of future probabilities (or, expectations).<sup>8</sup> Let  $\mathcal{P}(i, j) := \{Q_j | P_i = Q_i, P_i(\text{SP}_j(Q_j)) > 0\}$  be the range of probability measures at  $j$  in the branching trajectories that are subjectively possible from the point of view of  $i$ . I assume that  $\mathcal{P}(i, j)$  is at most countably infinite so that we can use sums instead of integrals:

$$\text{(Iteration)} \quad \forall i, j \ (i \leq j) : \\ E_{P_i}(X) = \sum_{Q_j \in \mathcal{P}(i, j)} P_i(\text{SP}_j(Q_j)) \times E_{Q_j}(X).$$

Van Fraassen (1995) suggested that present opinions should lie within the range of subjectively possible future opinions. If opinions are repre-

sented by probability functions, this translates into a further diachronic principle:<sup>9</sup>

(Generalized Reflection)  $\forall i, j (i \leq j) :$

$$\min_{Q_j \in \mathcal{P}(i,j)} Q_j(A) \leq P_i(A) \leq \max_{Q_j \in \mathcal{P}(i,j)} Q_j(A).$$

If opinions are more liberally understood as expectations, we obtain a third candidate:

(Generalized Iteration)  $\forall i, j (i \leq j) :$

$$\min_{Q_j \in \mathcal{P}(i,j)} E_{Q_j}(X) \leq E_{P_i}(X) \leq \max_{Q_j \in \mathcal{P}(i,j)} E_{Q_j}(X).$$

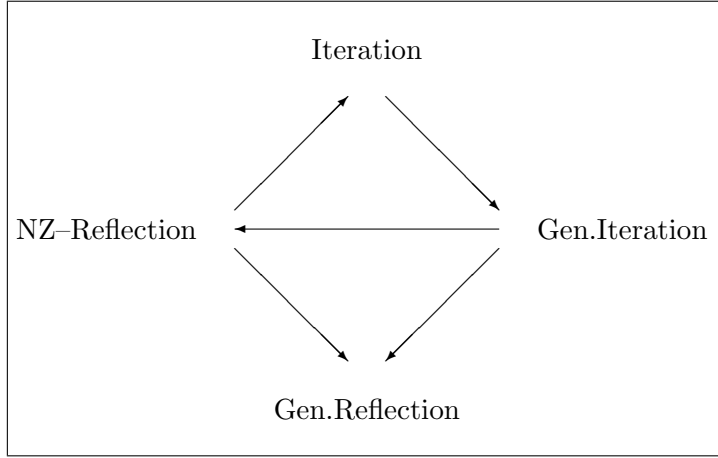
The logical relationships between these principles and (Non-Zero) Reflection are summarized in the following theorem.<sup>10</sup> Non-Zero Reflection and its equivalents, Iteration and Generalized Iteration, emerge as the central principles of diachronic auto-epistemology.

*Theorem 2.2 (AE-Principles).* (i) Non-Zero Reflection implies Iteration.

- (ii) Given AE-Transparency, Iteration implies Non-Zero Reflection.
- (iii) Non-Zero Reflection implies Generalized Reflection, but not conversely.
- (iv) Generalized Iteration implies Generalized Reflection, but not conversely.
- (v) Given AE-Transparency, Generalized Iteration implies Non-Zero Reflection.
- (vi) Iteration implies Generalized Iteration.

#### AUTO-EPISTEMIC UPDATE MODELS

We will now increase the auto-epistemic reasoner's capacities so that they include reasoning about update mechanisms. Let  $\mathcal{EV}$  be the set of (logically) possible pieces of total evidence.<sup>11</sup> An *update rule*  $\star : \mathcal{PR} \times \mathcal{EV} \rightarrow \mathcal{PR}$  at  $i$  maps prior probability measures and incoming total evidence from  $\mathcal{EV}$  into posterior probability measures. For simple Conditionalization  $\mathcal{EV}$  will simply be identical to  $\mathcal{A}^0$ . For Generalized Conditionalization, it will be the set of possible Jeffrey constraints  $\langle B_k, p_k \rangle_{k \in K}$  over  $\mathcal{A}^0$ .



*Legend:*  
 AE-Transparency is presupposed;  
 arrows indicate implication.

Figure 1. AE-Principles

An *evidence function*  $\pi : \Omega \times I \rightarrow \mathcal{EV}$  attaches at each stage  $i$  and to each world  $\omega$  the piece of total evidence that the reasoner will receive in  $\omega$  at  $i$ . Evidence functions presuppose and express a certain protocol according to which new evidence is presented, e.g. by sensory input, utterances made to the reasoner by other people, or by other channels of information. (I will later return to this point.) Let  $\text{Ev}_i^{\text{tot}}(e) := \{\omega \mid \pi_{\omega,i} = e\}$  be the event that the reasoner receives  $e$  as her total evidence at  $i$  ( $e \in \mathcal{EV}$ ). Let  $\mathcal{EV}(i, j) := \{e \mid P_i(\text{Ev}_j^{\text{tot}}(e)) > 0\}$  be the set of total  $j$ -evidence subjectively possible from the point of view of  $i$ . Let  $\star(\cdot)$  be a sequence of update rules for each  $i \in I$ . We then say that  $\langle \Omega, \mathcal{A}^{\text{AE}}, I, P_{(\cdot, \cdot)}, \star(\cdot), \pi \rangle$  is an *auto-epistemic update model* if and only if  $P_{(\cdot, \cdot)}$  is an epistemic function relative to  $\mathcal{A}^{\text{AE}}$  and  $\mathcal{A}^{\text{AE}}$  is the closure of  $\mathcal{A}^0$  under the auto-epistemic vocabulary  $\text{SP}_i(Q)$  and  $\text{Ev}_i^{\text{tot}}(e)$  (for all  $i \in I$ ,  $Q \in \mathcal{PR}$ , and  $e \in \mathcal{EV}$ ).

Updating in an auto-epistemic update model will comprise updating opinions about auto-epistemic propositions such as ‘ $e$  is the total evidence at  $i$ ’ ( $\text{Ev}_i^{\text{tot}}(e)$ ). A special class of update functions is evidentially transparent in the sense that the grounds for the reasoner to change

her opinion are exactly what she (after the change) believes to be her total evidence:

$$\text{(EV-Transparency)} \quad [P_i^{\star_{i+1}} e](\text{Ev}_{i+1}^{\text{tot}}(e')) = \begin{cases} 1, & \text{if } e = e'. \\ 0, & \text{otherwise.} \end{cases}$$

Since evidentially transparent update functions state exactly what evidence has been received, updates with different pieces of evidence will lead to different posterior probability measures (hence, each  $\star_i$  is an injection).

We say that an auto-epistemic update model is *evidence driven* if and only if changes in probabilities exclusively result from updates with incoming evidence (i.e.,  $P_{\omega, i+1} = P_{\omega, i}^{\star_{i+1}} \pi(\omega, i+1)$ ). The next theorem states that, in evidence driven models, Reflection takes the form of the important principle of *Auto-Epistemic Conditionalization*:

$$\text{(AE-Cond)} \quad P_i(\cdot)^{\star_{i+1}} e = P_i(\cdot | \text{Ev}_{i+1}^{\text{tot}}(e))$$

*Theorem 2.3.* In evidence driven auto-epistemic update models, Reflection is equivalent to AE-Conditionalization.

### 3. The Impact of Auto-Epistemology on Updating

#### AE-CONDITIONALIZATION AND UPDATES

AE-Conditionalization operates on a level very different from (Generalized) Conditionalization. The difference is the same as that between auto-epistemology on the one hand, and methodology on the other. AE-Conditionalization requires that your update methods be taken into account by your current probabilities conditional on statements about future evidence, *whatever* these update methods may be. In a nutshell, AE-Conditionalization commands: ‘Know how you update!’ (Generalized) Conditionalization, on the other hand, specifies what universal update method you should use: ‘When presented with new information  $B$ , use your old probabilities conditional on  $B$  as your updated probabilities!’

Auto-epistemology and universal methodologies will often clash because a reasoner in an auto-epistemic update model can assess her learning situation prior to choosing her update methodology for this particular situation. Depending on her assessment, AE-Conditionalization then implies an update method that is individually

suited to this particular situation but not universally applicable. What assessment of the learning situation will lead to the use of Conditionalization? Put differently, under what auto-epistemic conditions does AE-Conditionalization imply Conditionalization? For simple Conditionalization, this question is easily answered by the following equation. It gives a necessary as well as sufficient condition for the equivalence of AE-Conditionalization and Conditionalization ( $i \leq j$ ):

$$\text{(Crux)} \quad P_i(A|\text{Ev}_{i+1}^{\text{tot}}(B)) = P_i(A|B), \quad A, B \in \mathcal{A}^0.^{12}$$

To gain a more detailed insight into the relationship between AE-Conditionalization and (Generalized) Conditionalization, we can divide this equation into a Reliability and an Evidential Independence condition:<sup>13</sup>

*Theorem 3.4.* Let  $e_{GC} := \langle B_k, p_k \rangle_{k \in K}$  and  $A \in \mathcal{A}^0$ . Given AE-Conditionalization, Success is equivalent to belief in the reliability of evidential input:

$$\text{(Reliability)} \quad P_i(B_k|\text{Ev}_{i+1}^{\text{tot}}(e_{GC})) = p_k$$

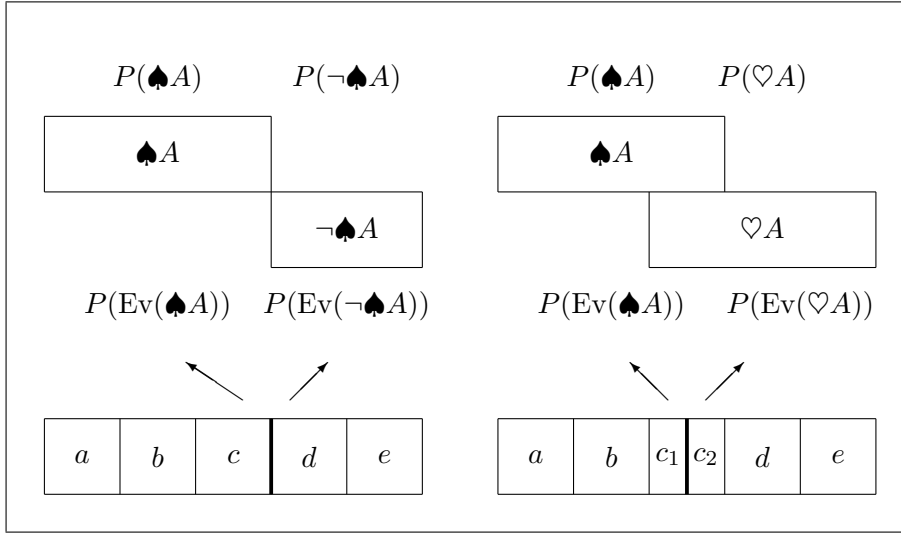
Given AE-Conditionalization, Rigidity is equivalent to the probabilistic independence of  $A$  from  $\text{Ev}_{i+1}^{\text{tot}}(e_{GC})$  given  $B_k$ :

$$\text{(Ev.Independence)} \quad P_i(A|B_k \cap \text{Ev}_{i+1}^{\text{tot}}(e_{GC})) = P_i(A|B_k)$$

AE-Conditionalization and (Generalized) Conditionalization coincide if and only if Reliability and Evidential Independence hold.

Remember that (Generalized) Conditionalization is equivalent to Success plus Rigidity. Reliability and Evidential Independence are the auto-epistemic correlates of these two conditions. For this reason, AE-Conditionalization validates (Generalized) Conditionalization if and only if the auto-epistemic conditions of Reliability and Evidential Independence hold. AE-Conditionalization outlaws (Generalized) Conditionalization if and only if Reliability and Evidential Independence fail.

Textbooks on probability theory normally consider only examples in which Reliability and Evidential Independence are implicitly assumed to hold. Such examples are chance experiments about which the reasoner has the following full beliefs: Firstly, she fully believes in the reliability of her evidence. Secondly, she fully believes that the possible outcomes of the experiment are mutually exclusive and jointly exhaustive of the space of all subjective possibilities. She thus believes that her possible future evidence forms a partition  $\{B_k|k \in K\}$  of the support



*Legend:*

Only worlds with non-zero probability are shown.  $\spadesuit A = \{a, b, c\}$ ,  $\heartsuit A = \{c, d, e\}$ .

Left half:  $\text{Ev}_{i+1}^{\text{tot}}(\spadesuit A) = \{a, b, c\}$ ,  $\text{Ev}_{i+1}^{\text{tot}}(\neg\spadesuit A) = \{d, e\}$ .

Right half:  $\text{Ev}_{i+1}^{\text{tot}}(\spadesuit A) = \{a, b, c_1\}$ ,  $\text{Ev}_{i+1}^{\text{tot}}(\heartsuit A) = \{c_2, d, e\}$ .

Figure 2. Assessments of Evidence

of  $P_i(\cdot)$ .<sup>14</sup> We say that the reasoner makes the Partitioning assumption if and only if the following holds:

(Partitioning)            Let  $\mathcal{EV}(i, i+1) = \{B_k | k \in K\}$ . Then

- (i)      $P_i(B_k \cap \neg B_{k'}) = 0$                        $(k \neq k')$ ,
- (ii)     $P_i(\bigcup_{k \in K} B_k) = 1$ ,                       $(k \in K)$ .

Learning situations of this sort are sketched in the left half of Figure 14. (The particular labels of the events refer to Freund's Puzzle in Section 4.) It turns out that this assessment of her learning situation satisfies Evidential Independence:

*Theorem 3.5.* Partitioning and Reliability imply Evidential Independence.

AE-Conditionalization therefore leads a reasoner who believes in Reliability and makes the Partitioning assumption to update by Conditionalization on the outcome  $B_k$  of the experiment.

## RELIABILITY AND EVIDENTIAL INDEPENDENCE

Let us now consider learning situations in which AE-Conditionalization is at odds with (Generalized) Conditionalization. Receiving total evidence  $B$ , as I understand it, means to be presented (precisely) with the information that  $B$  has occurred, or to receive (precisely) the input that  $B$ . Before we can unambiguously describe a learning situation or an update methodology, we must specify how evidence, or input, will be provided, e.g. by sensorial input, by sentences that appear on a screen, or by sentences that are uttered to the reasoner by other people. Once we have fixed this protocol, it must not be changed again. In auto-epistemic update models, such protocols enter into the reasoner's evidence function  $\pi$  and thus into her assessment of her learning situation. When  $\pi$  takes the value  $e$ , this means that  $e$  is presented as total input relative to the protocol.

Objections against the choice of any particular protocol may be raised only before the applicability of an update method is discussed. On pains of trivializing Conditionalization, its champions must not retrospectively manipulate a fixed protocol  $\pi$  after (and because) a violation of Conditionalization has been detected. For it is in many cases possible to construct a manipulated protocol  $\pi^*$  that makes a given transition from prior to posterior measure trivially satisfy Conditionalization. (I will return to this point later.) The notion of the total evidence (to be taken as input by an update rule) is therefore always relative to a given protocol. The following counter-examples to Conditionalization construct learning situations and subjective assessments that violate Reliability or Evidential Independence. The same type of counter-examples can be constructed no matter how the initial protocol is chosen.

**(Example 3.1)** Assume that the reasoner takes utterances made to her as evidence. Consider the extreme case in which the reasoner is certain that somebody who tells her that  $A$  will try to deceive her:  $P_i(-A|\text{Ev}_{i+1}^{\text{tot}}(A)) = 1$ . Conditionalization would thus have her adopt the posterior probability measure  $P_i(\cdot|A)$ . She is, however, certain that  $-A$  when she receives the input  $A$ . She will therefore assign maximal probability to  $-A$  and violate Success.

**(Example 3.2)** Assume that the reasoner takes the readings of a measuring instrument as evidence. She knows that a measuring instrument is a good, but not perfect, indicator for  $A$  and to a lesser degree a good indicator for  $-A$ :

$$\begin{array}{ll} P_i(A|\text{Ev}_{i+1}^{\text{tot}}(A)) = .9 & P_i(-A|\text{Ev}_{i+1}^{\text{tot}}(A)) = .1 \\ P_i(A|\text{Ev}_{i+1}^{\text{tot}}(-A)) = .2 & P_i(-A|\text{Ev}_{i+1}^{\text{tot}}(-A)) = .8 \end{array}$$

Upon receiving  $A$  as total evidence, she would then only assign a probability of .9 to  $A$  and thus violate Success. In the absence of auto-epistemic vocabulary, we could only capture the reasoner's updating method by changing the protocol and by re-describing the evidence as  $\{\langle A, .9 \rangle, \langle -A, .1 \rangle\}$  if the instrument outputs the statement ' $A$ ' and as  $\{\langle A, .2 \rangle, \langle -A, .8 \rangle\}$  if it outputs ' $-A$ '.<sup>15</sup>

Evidential Independence (and hence, Rigidity) can be violated even if evidence is subjectively reliable.

**(Example 3.3)** Consider a police officer whose evidential input consists of statements made by a witness. Her probability for apprehending Mr. X ( $A$ ) given that he has committed a burglary ( $B$ ) is less than her probability of apprehending him given that his burglary has been observed and is reported by a witness ( $Ev_{i+1}^{tot}(B)$ ). Hence, the fact that she receives the evidence (the witness's report) that Mr. X has committed a burglary changes her probability for his arrest conditional on him being a burglar. Hence, Evidential Independence fails and the reasoner does not update by means of Conditionalization on  $B$ .

Assessments of learning situations in which Evidential Independence fails are sketched in the right half of Figure 14. In Section 4, I will *in extenso* discuss Freund's Puzzle as a paradigmatic example of such situations.

#### CONDITIONALIZING ON EVIDENCE

These examples show how easy it is to find auto-epistemic assessments of learning situations that force the reasoner to violate Conditionalization for the sake of AE-Conditionalization. Conversely, a Conditionalizer does not automatically satisfy AE-Conditionalization. All depends on her subjective assessment of her learning situation. Let me summarize:

- Observation 3.6.* (i) AE-Conditionalization and (Generalized) Conditionalization are logically independent.
- (ii) AE-Conditionalization is incompatible with (Generalized) Conditionalization as a universal update rule.

In the literature, the auto-epistemic principle of AE-Conditionalization is often identified with the update method of Conditionalization (cf. Skyrms (1980), Jeffrey (1988)). In my view, this results from a laxness to specify all elements of the proposed probabilistic model explicitly.

This laxness surfaces, for example, in the confusion surrounding Freund's Puzzle and similar scenarios. In particular, the underlying protocol (evidence function)  $\pi$  is often not made explicit. Because the notion of total evidence is relative to a protocol, it becomes unclear what information exactly the reasoner receives. The distinction between the proposition  $B$  and the proposition that  $B$  is received as evidence is therefore easily overlooked. If, according to the evidence function and the underlying protocol, the reasoner receives input  $B$  as total evidence, Conditionalization prescribes that she should use  $P_i(\cdot|B)$  as her posterior probability measure  $P_i^*B$ . This differs from conditionalizing on the proposition  $\text{Ev}_{i+1}^{\text{tot}}(B) = \{\omega|\pi(\omega, j) = B\}$  that she receives  $B$  as evidence.

Advocates of Conditionalization must be careful not to trivialize the content of their update methodology. In each learning situation, they must explicitly specify the protocol and the notion of total evidence for which they want to propose updates by Conditionalization. Otherwise, they are in danger of begging the question. Suppose we are given a prior probability measure  $P$  and an updated probability measure  $Q$  that violate Conditionalization relative to some protocol and piece of total evidence  $E$ . Retrospectively, it is in many cases possible to manipulate the protocol and the notion of total evidence in such a way that the updated measure equals the prior measure conditional on some piece of pseudo-evidence  $E^*$  thus manipulated.

Suppose we have a probability space  $\langle \Omega, \mathcal{A} \rangle$  with a prior probability measure  $P$ , and an updated measure  $Q = P^*E$  ( $E \in \mathcal{A}$ ). Suppose furthermore that  $Q$  is not updated by Conditionalization of  $P$  on  $E$  ( $Q(\cdot) \neq P(\cdot|E)$ ). The below theorem states in which cases we can construct pseudo-evidence  $E^*$  such that  $Q$  can be obtained from Conditionalization of  $P$  on  $E^*$ . A probability space  $\langle \Omega, \mathcal{A} \rangle$  with measures  $P$  and  $Q$  can be embedded into a probability space  $\langle \Omega^*, \mathcal{A}^* \rangle$  with measures  $P^*$  and  $Q^*$  if and only if there is a homomorphism  $\phi : \mathcal{A} \rightarrow \mathcal{A}^*$  such that  $P(A) = P^*(\phi(A))$  and  $Q(A) = Q^*(\phi(A))$  (for all  $A \in \mathcal{A}$ ). We say that *Conditionalization can be trivialized for  $P$  and  $Q$*  if and only if the probability space  $\langle \Omega, \mathcal{A} \rangle$  can be embedded into a probability space  $\langle \Omega^*, \mathcal{A}^* \rangle$  with measures  $P^*$  and  $Q^*$  such that there is an event  $E^* \in \mathcal{A}^*$  with  $Q^*(\phi(A)) = P^*(\phi(A)|E^*)$  for all events  $A \in \mathcal{A}$  and  $P^*(E^*) > 0$ .

*Theorem 3.7* (Diaconis/Zabell (1982)). Conditionalization can be trivialized for probability measures  $P$  and  $Q$  over  $\langle \Omega, \mathcal{A} \rangle$  if and only if  $\exists b \geq 1$  such that for all  $\omega \in \Omega$ :  $Q(\omega) \leq bP(\omega)$ .

#### 4. Application: Freund's Puzzle

Let us take a more detailed look at the methodological impact of auto-epistemic reasoning in the analysis of the learning situation constructed in Freund's (1965) Puzzle.<sup>16</sup> I will argue that differences in the auto-epistemic assessment of the learning situation resolve the methodological ambiguities of the puzzle. On the background of different assessments, auto-epistemic reasoning prescribes different methodological reactions to the incoming evidence. These background assumptions always seem to contain Reliability, but satisfy Evidential Independence in only one, extreme, case. In that particular case, updating must proceed by Conditionalization. In all other cases, AE-Conditionalization will prescribe updates which violate Conditionalization. This analysis agrees with Shafer (1985), but differs from Jeffrey's (1988) solution of the related Three Prisoners Puzzle.<sup>17</sup>

Two cards are drawn at random from a four-card deck containing  $\spadesuit A$ ,  $\spadesuit 2$ ,  $\heartsuit A$ ,  $\heartsuit 2$ . You receive evidence about this random experiment in the form of utterances made to you by the game master. You are then told that one of the two drawn cards is an ace. This leaves you with a uniform probability distribution  $P_1(\cdot)$  over the remaining possibilities:<sup>18</sup>

$$\begin{aligned} P_1(a) &= P_1(\spadesuit A \wedge \heartsuit A) = \frac{1}{5} \\ P_1(b) &= P_1(\spadesuit A \wedge \spadesuit 2) = \frac{1}{5} \\ P_1(c) &= P_1(\spadesuit A \wedge \heartsuit 2) = \frac{1}{5} \\ P_1(d) &= P_1(\heartsuit A \wedge \spadesuit 2) = \frac{1}{5} \\ P_1(e) &= P_1(\heartsuit A \wedge \heartsuit 2) = \frac{1}{5} \end{aligned} \quad (2)$$

(The labels 'a', 'b' etc. refer to Figure 14.) In particular, the probability for both aces having been drawn is  $\frac{1}{5}$ . Now consider the additional utterance (information) that one of the drawn cards is the ace of spades. You believe that what you are told is reliable:

$$P_1(\spadesuit A | \text{Ev}_2^{\text{tot}}(\spadesuit A)) = 1 \quad (3)$$

You consequently want your update function to incorporate the new information you just received (Success). How should you update your probabilities? The puzzle turns out to play on an ambiguity in how to understand the information that the ace of spades has been drawn. People typically tend to respond to this information in two specific ways that we will now re-construct as different auto-epistemic assessments of this learning situation. If you find neither assessment intuitively appealing, simply jump to the disambiguated general solution.

**Assessment 1:** You believe that you will be told correctly whether or not the ace of spades has been drawn:

$$P_1(\text{Ev}_2^{\text{tot}}(\spadesuit A) | \spadesuit A) = 1$$

$$P_1(\text{Ev}_2^{\text{tot}}(\neg\spadesuit A)|\neg\spadesuit A) = 1$$

This assessment satisfies Evidential Independence. It is illustrated in the left half of Figure 14. The information that the ace of spades has been drawn therefore means to you that the last two options on list (2) have been eliminated. Since you have gained no extra information about any one of the remaining three cases, you assume that they are equiprobable ( $\frac{1}{3}$ ). In other words, you update by Conditionalization on  $\spadesuit A$ .

**Assessment 2:** You believe that, if only one ace has been drawn, you will be told the colour of that ace:

$$\begin{aligned} P_1(\text{Ev}_2^{\text{tot}}(\spadesuit A)|\spadesuit A \wedge \neg\heartsuit A) &= 1 \\ P_1(\text{Ev}_2^{\text{tot}}(\heartsuit A)|\heartsuit A \wedge \neg\spadesuit A) &= 1 \end{aligned}$$

You also believe that the flip of a fair coin decides which colour is announced to you, should both aces have been drawn:

$$P_1(\text{Ev}_2^{\text{tot}}(\spadesuit A)|\spadesuit A \wedge \heartsuit A) = \frac{1}{2}$$

This assessment violates Evidential Independence. It is illustrated in the right half of Figure 14. Since you know already that at least one ace has been drawn, the announcement only tells you the colour of a drawn ace. By learning its colour, you have learnt nothing about the second ace and nothing about whether or not both aces have been drawn. You should therefore stick to your old probability of  $\frac{1}{5}$  for both aces having been drawn. In other words, you defy Conditionalization.

**General Solution:** You believe (3) that the colour that is announced to you is indeed the colour of a drawn ace (Reliability). You also believe that the spadesuit is correctly announced unless there is a choice to be made:

$$P_1(\text{Ev}_2^{\text{tot}}(\spadesuit A)|\spadesuit A \wedge \neg\heartsuit A) = 1 \quad (4)$$

You finally believe that, in case of a choice (i.e., if both aces have been drawn), the spadesuit is announced to you with probability  $p$ :

$$P_1(\text{Ev}_2^{\text{tot}}(\spadesuit A)|\spadesuit A \wedge \heartsuit A) = p \quad (5)$$

Obviously, Assessment 1 and Assessment 2 are special instances of equation (5) with  $p = 1$  and  $p = \frac{1}{2}$ , respectively. The two assumptions (3) and (4) are common to Assessment 1 and Assessment 2. If they were given up (as might be plausible), the following calculations would become slightly more complicated but would still illustrate the

same point, namely that, for the learning situation described in the puzzle, Conditionalization is only appropriate under exceptional circumstances.

Your auto-epistemic assessment is now detailed enough uniquely to prescribe an update method and posterior probabilities. According to AE-Conditionalization, your updated probabilities equal your prior probabilities conditional on the proposition that  $\spadesuit A$  is your total evidence ( $P_1(\cdot | \text{Ev}_2^{\text{tot}}(\spadesuit A))$ ). Your opinion implies a probability for being told that the ace of spades has been drawn:

$$P_1(\text{Ev}_2^{\text{tot}}(\spadesuit A)) = \frac{1}{5}p + \frac{2}{5}$$

Further calculations show that your updated probabilities depend on  $p$  in the following manner:

	$P_1(\cdot)$	$P_2(\cdot)$	$p = 1$	$p = \frac{1}{2}$
$\spadesuit A \wedge \heartsuit A$	$\frac{1}{5}$	$\frac{p}{p+2}$	$\frac{1}{3}$	$\frac{1}{5}$
$\spadesuit A \wedge \spadesuit 2$	$\frac{1}{5}$	$\frac{1}{p+2}$	$\frac{1}{3}$	$\frac{2}{5}$
$\spadesuit A \wedge \heartsuit 2$	$\frac{1}{5}$	$\frac{1}{p+2}$	$\frac{1}{3}$	$\frac{2}{5}$
$\heartsuit A \wedge \spadesuit 2$	$\frac{1}{5}$	0	0	0
$\heartsuit A \wedge \heartsuit 2$	$\frac{1}{5}$	0	0	0

Assessment 1 and 2 are included in the columns under  $p = 1$  and  $p = \frac{1}{2}$ . An inspection of the table teaches that the update proceeds by Conditionalization only when  $p = 1$ . (Only  $p = 1$  fulfils Evidential Independence.) For all other values of  $p$ , your assessment of your learning situation implies a violation of Conditionalization.

## 5. Propositional Auto-Epistemology

### AUTO-EPISTEMIC MODELS

Let us now turn to the auto-epistemology and updating of full belief. I will begin with a propositional characterization of auto-epistemic models in terms of possible worlds. As before, let  $\Omega$  be a non-empty set of possible worlds,  $\mathcal{A}^0$  some  $\sigma$ -algebra over  $\Omega$  that we are going to enrich with auto-epistemic vocabulary. For any  $\sigma$ -algebra  $\mathcal{A}$ ,  $\mathbf{K}$  is a *propositional belief set* over  $\mathcal{A}$  if and only if  $\emptyset \neq \mathbf{K} \in \mathcal{A}$ . ( $\mathcal{A}$  puts restrictions on the expressiveness of the reasoner's concepts and his ability to discriminate between different possible worlds.) A reasoner is said to believe a proposition  $\mathbf{H} \in \mathcal{A}$  in  $\omega$  at  $i$  if and only if  $\mathbf{K}_{\omega,i} \subseteq \mathbf{H}$ .

Let  $I$  be a diachronic index. An *epistemic function* relative to  $\mathcal{A}$  is a function  $\mathbf{K}_{(\cdot, \cdot)} : \Omega \times I \rightarrow \mathcal{A}$  that attaches a propositional belief set to each possible world  $\omega$  and each stage  $i$ . Let  $\mathbf{B}_i(\mathbf{H}) := \{\omega \mid \mathbf{K}_{(\omega, i)} = \mathbf{H}\}$  be the event that the reasoner's belief set at  $i$  is  $\mathbf{H}$  (where  $\mathbf{H} \in \mathcal{A}$ ) and  $\mathbf{B}_i(\mathbf{A}) := \{\omega \mid \mathbf{K}_{\omega, i} \subseteq \mathbf{A}\}$  be the event that the reasoner believes  $\mathbf{A}$  at  $i$  (where  $\mathbf{A} \in \mathcal{A}$ ).

Evidence functions  $\pi : \Omega \times I \rightarrow \mathcal{EV}$  and the event  $\text{Ev}_i^{\text{tot}}(e) := \{\omega \mid \pi_{\omega, i} = e\}$  are defined as before. For  $i \in I$ , an *update rule*  $\star_i : \mathcal{A} \times \mathcal{EV} \rightarrow \mathcal{A}$  maps prior belief sets and incoming evidence into posterior belief sets. We then say that  $\langle \Omega, \mathcal{A}^{AE}, I, \mathbf{K}_{(\cdot, \cdot)}, \star(\cdot), \pi \rangle$  is an *auto-epistemic update model* if and only if  $\mathbf{K}_{(\cdot, \cdot)}$  is an epistemic function relative to  $\mathcal{A}^{AE}$  and  $\mathcal{A}^{AE}$  is the closure of  $\mathcal{A}^0$  under the auto-epistemic vocabulary  $\mathbf{B}_i(\mathbf{K})$  and  $\text{Ev}_i^{\text{tot}}(e)$  (for all  $i \in I$ ,  $\mathbf{K} \in \mathcal{A}^{AE}$ , and  $e \in \mathcal{EV}$ ).

#### SENTENTIAL VS. PROPOSITIONAL MODELS

In the literature, models of full belief are often characterized sententially in terms of a formal language  $\mathcal{L}$  and a consequence operator  $\text{Cn} : 2^{\mathcal{L}} \rightarrow 2^{\mathcal{L}}$ .<sup>19</sup> *Sentential belief sets* are consistent (i.e., proper) subsets of  $\mathcal{L}$  that are closed under  $\text{Cn}(\cdot)$ . (The ‘absurd belief set’  $\mathcal{L}$  is not a belief set in this sense.) Full belief in a sentence  $A$  then corresponds to membership of  $A$  in  $K_i$ , dis-belief to the membership of its negation, and non-belief to non-membership. To examine the relationship between propositional and sentential models, let us consider a synchronic propositional model  $\langle \Omega, \mathcal{A}, \mathbf{K} \rangle$  of full belief (where  $\mathbf{K}$  is a propositional belief set), and a synchronic sentential model  $\langle \mathcal{L}, \text{Cn}, K \rangle$  of full belief (where  $K$  is a sentential belief set). An *interpretation*  $[\cdot] : \mathcal{L} \rightarrow \mathcal{A}$  relates each sentence in  $\mathcal{L}$  to an event in  $\mathcal{A}$ . Interpretations must respect the logical structure of  $\mathcal{L}$  in the usual way ( $[A \wedge B] = [A] \cap [B]$ ,  $[\neg A] = -[A]$ , etc.). For a set of sentences  $\Gamma \subseteq \mathcal{L}$ , we define  $[\Gamma] := \bigcap_{A \in \Gamma} [A]$ . (This definition extends the interpretation  $[\cdot]$  it into a function from  $2^{\mathcal{L}}$  to  $\mathcal{A}$ .)

Sentential and propositional models of full belief deliver an essentially identical description of a reasoner as long as  $\mathcal{A}$  and  $\mathcal{L}$  endow him with the same conceptual expressiveness. By this I mean that the following two conditions must be satisfied:

(Expressiveness 1) For every event  $\mathbf{E} \in \mathcal{A}$  there must be a set of sentences  $\Gamma \subseteq \mathcal{L}$  such that  $\mathbf{E} = [\Gamma]$ .

(Expressiveness 2)  $A \in \text{Cn}(\Gamma)$  if and only if  $[\Gamma] \subseteq [A]$ .

These conditions imply that every propositional belief set can be represented by a sentential belief set.<sup>20</sup>

I will follow the conventions of the literature and discuss models of full belief from a sentential point of view because I believe that this will make the discussion more intuitive for most readers. If we want to introduce auto-epistemic models purely sententially, however, we have to provide suitable axioms that govern the Cn-logic of auto-epistemic vocabulary. We also need to supply extensionality axioms that require logically equivalent sentences to be treated equally, since a sentential model can distinguish between syntactically different but logically equivalent  $\mathcal{L}$ -descriptions of the same  $\mathcal{A}$ -event. Since this is a trivial but tedious business, I will use the above propositional definition of auto-epistemic update models as basic and assume that Expressiveness 1 and 2 hold for  $\mathcal{L}^{AE}$  and  $\mathcal{A}^{AE}$ . We can then define  $K_{(\omega,i)}$  to be the sentential belief set that corresponds to  $\mathbf{K}_{(\omega,i)}$  (i.e.,  $[K_{(\omega,i)}] = \mathbf{K}_{(\omega,i)}$ ). Similarly, for  $\Gamma \subseteq \mathcal{L}^{AE}$ ,  $B_i(\Gamma)$  be a sentence from  $\mathcal{L}^{AE}$  such that  $[B_i(\Gamma)] = \mathbf{B}_i([\Gamma])$ . As in the probabilistic case, we focus on two trajectories  $K_i := K_{(\omega_1,i)}$  and  $H_i := K_{(\omega_2,i)}$  ( $i \in I$ ).

#### CONDITIONAL BELIEFS

Since the structure of full beliefs corresponds to that of maximal probabilities, we can formally base belief sets on probability functions. Let  $K_{Q(\cdot)} := \{A \in \mathcal{L} \mid Q([A]) = 1\}$ .<sup>21</sup>

$$\text{(Correlation 1)} \quad K_i = K_{P_i(\cdot)}$$

The probability axioms then imply that  $K_i$  is a sentential belief set. As a matter of fact, the following presentation of propositional auto-epistemology and updating can entirely be derived from the probabilistic case under Correlation 1 and Correlation 2 below. Despite this formal correspondence, it is, of course, a different question whether or not the interpretation of full beliefs should coincide with that of extreme probabilities.

Correlation 1 can be extended to include conditional beliefs  $K_i|A$ :

$$\text{(Correlation 2)} \quad K_i|A = K_{P_i(\cdot|A)}$$

We will see below that ‘conditional beliefs’ as characterized by Correlation 2 correspond exactly to ‘revisions’ in the AGM framework (cf. Gärdenfors (1988)). According to the AGM framework, a revision of a belief set  $K_i$  by  $A$  should result in a consistent belief set that contains  $A$  and requires ‘minimal changes’ of  $K_i$  (necessary for preserving consistency). If  $A$  is compatible with  $K_i$  (i.e., if  $K_i$  does not contain  $\neg A$ ), a revision of  $K_i$  by  $A$  should therefore result in adding merely  $A$  and its

consequences to  $K_i$ . The operation of adding  $A$  and its consequences to  $K_i$  is called ‘expansion’:

$$\text{(Expansion)} \quad K_i^+ A := \text{Cn}(K_i \cup \{A\})$$

$$\text{Observation 5.8. } B \in K^+ A \quad \text{iff} \quad (A \rightarrow B) \in K.$$

If input  $A$  is incompatible with  $K_i$  (because  $\neg A$  is contained in  $K_i$ ), the expansion of  $K_i$  by  $A$  would result in the ‘absurd belief set’  $\mathcal{L}$ . A full-blown revision operator must therefore be more elaborate than Expansion. The following AGM axioms characterize some of its properties:<sup>22</sup>

- (K|1)  $K_i|A = \text{Cn}(K_i|A)$ .  
(Closure)
- (K|2)  $A \in K_i|A$ , unless  $A \iff \perp$ .  
(Success)
- (K|3)  $K_i|A \subseteq K_i^+ A$ .  
(Expansion 1, Inclusion)
- (K|4) If  $\neg A \notin K_i$ , then  $K_i^+ A \subseteq K_i|A$ .  
(Expansion 2)
- (K|5)  $\perp \notin K_i|A$ , unless  $A \iff \perp$ .  
(Consistency)
- (K|6) If  $A \iff B$ , then  $K_i|A = K_i|B$ .  
(Extensionality)
- (K|7)  $K_i|(A \wedge B) \subseteq (K_i|A)^+ B$ .  
(Conjunction 1)
- (K|8) If  $\neg B \notin K_i|A$ , then  $(K_i|A)^+ B \subseteq K_i|(A \wedge B)$ .  
(Conjunction 2, Rational Monotony)

Under Correlation 2, we obtain the AGM axioms for revisions from the axioms (CP1)–(CP4) for conditional probabilities. It is easy to see that (CP1) yields (K|1) and (K|5), that (CP2) yields (K|2), and that Expressiveness 2 yields Extensionality (K|6).<sup>23</sup> Multiplication (CP3) and Reduction (CP4) translate into:<sup>24</sup>

- (Multiplication<sub>K</sub>)  
 $B \in K_i|C$  and  $A \in K_i|(B \wedge C)$  iff  $(A \wedge B) \in K_i|C$   
 $\neg B \notin K_i|C$  and  $A \notin K_i|(B \wedge C)$  iff  $(B \rightarrow A) \notin K_i|C$
- (Reduction<sub>K</sub>)  $K_i|\top = K_i$

*Observation 5.9.* Given  $(K|1)$ , the first line of  $\text{Multiplication}_K$  is redundant (follows from the second line).

*Theorem 5.10.* Given  $(K|1)$  and  $(K|6)$ ,  $\text{Multiplication}_K$  and  $\text{Reduction}_K$  are equivalent to  $(K|3)$ ,  $(K|4)$ ,  $(K|7)$ , and  $(K|8)$ .

The last theorem establishes that under Correlation 2,  $(\text{CP1})$ – $(\text{CP4})$  correspond to  $(K|1)$ – $(K|8)$ . The AGM axioms for revisions of full belief are hence equivalent to the axiomatization of conditional full beliefs.

#### AGM UPDATING

The AGM axioms are commonly presented as axioms that specify the logical properties of update methods. In the same spirit in which Conditionalization is put forward as a universal probabilistic update rule, AGM revisions are often looked upon as universal update procedure. AGM–Updating claims that the same axioms that hold for conditional beliefs also hold for updating, or, equivalently, that updating proceeds by shifting to conditional prior beliefs:

$$(\text{AGM–Updating}) \quad K_i^* A = K_i | A$$

In order to avoid confusion, I shall reserve the axioms  $(K|1)$  –  $(K|8)$  to conditional belief only (revisions in the narrow sense). When the AGM axioms are interpreted as update principles, I shall refer to them as ‘ $(K^*1)$  –  $(K^*8)$ ’. Below, I will argue that conditional beliefs (AGM revised beliefs) are not *per se* identical to updated beliefs and that their identification constitutes a substantial axiom in itself.

#### SYNCHRONIC AUTO-EPISTEMOLOGY

AE–Transparency requires that a belief set contains  $A$  if and only if it contains the proposition that  $A$  is believed:<sup>25</sup>

$$(\text{AE–Transparency}_K) \quad \begin{array}{l} A \in K_i \text{ iff } B_i(A) \in K_i \\ A \notin K_i \text{ iff } \neg B_i(A) \in K_i \end{array}$$

AE–Transparency $_K$  has the immediate consequence that Moore–paradoxical statements cannot be believed:

$$(\text{Anti–Moore}) \quad (A \wedge B_i(\neg A)) \notin K_i$$

In other words, the reasoner must always believe in her cognitive infallibility. This consequence highlights the extent to which the present model idealizes rational belief. While the reasoner may think of other reasoners that they mistakenly believe  $A$ , Anti–Moore forbids any

doubts about his own cognitive integrity. Put differently, a state of ideal belief is no sooner attained than all doubts about its integrity are dissolved.

#### DIACHRONIC AUTO-EPISTEMOLOGY

Imagine that you are certain that you will believe  $A$  tomorrow. Then you should already believe  $A$  today since you would otherwise adopt an opinion of which you are certain that it will be outdated tomorrow. On the other hand, if you believe  $A$ , then you must not be certain that you will disbelieve it tomorrow no matter what. If your beliefs are to have any value for handling future events, you should not content yourself with such ephemeral opinions (cf. example on page 6). We can formally express these two conditions in terms of the range of belief sets at  $j$  that are subjectively possible from the point of view of  $i$ . Let  $\mathcal{K}(i, j) := \{H_j | H_i = K_i, \neg B_j(H_j) \notin K_i\}$ . Your present beliefs must lie within the subjectively possible range of your future beliefs:

(Generalized Reflection $_K$ )  $\forall i, j (i \leq j) :$

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

For similar reasons, you should only adopt an opinion of which you are certain that you will still entertain it tomorrow. Otherwise, you would already today countenance the possibility that your belief in  $A$  will be unwarranted tomorrow. This would be the diachronic version of a synchronic Moore incoherence (believing, firstly, that you do not believe that  $A$  but also, secondly, that  $A$ ). For ideal believers, Generalized Reflection can therefore be strengthened further:

(Iteration $_K$ )  $\forall i, j (i \leq j) :$

$$K_i = \bigcap \mathcal{K}(i, j)$$

Iteration $_K$  does not require that a belief should never be given up again, but rather that, in an ideal belief state, this should be expected not to happen. It rules out the subjective possibility of future belief sets in which present beliefs are disconfirmed and retracted.

Consider, finally, the analogue of Reflection in the context of full belief:

(Reflection $_K$ )  $K_i | B_j(H_j) = H_j, \quad i \leq j, K_i = H_i.$

Let Expansive Reflection $_K$  be the restriction of Reflection $_K$  to cases where  $H_j \in \mathcal{K}(i, j)$  and hence  $K_i | B_j(H_j) = K_i^+ B_j(H_j)$ .

*Theorem 5.11.* (i) Expansive Reflection $_K$  implies Iteration $_K$ .

- (ii)  $\text{Iteration}_K$  implies  $\text{Expansive Reflection}_K$ .
- (iii)  $\text{Iteration}_K$  implies  $\text{Generalized Reflection}_K$ , but not *vice versa*.

*Observation 5.12.* Under Correlation 1 and 2,  $\text{Generalized Reflection}_K$ ,  $\text{Iteration}_K$ ,  $\text{Expansive Reflection}_K$ , and  $\text{Reflection}_K$  correspond to Generalized Reflection, Iteration, Non-Zero Reflection, and Reflection, respectively.

#### APPLICATION: SURPRISE EXAMINATION

Wright/Sudbury (1977) suggest that the surprise examination paradox constitutes a counter-example to  $\text{Iteration}_K$ , although Binkley (1968) has already pointed out that the paradox ‘reduces to the phenomenon of incredible though possibly true propositions’.<sup>26</sup>  $\text{Iteration}_K$  is therefore as paradoxical for the diachronic case as is Anti-Moore for the synchronic. In other words,  $\text{Iteration}_K$  idealizes the diachronic aspect of rational belief in the same way in which AE-Transparency and Anti-Moore idealize its synchronic structure. It is also worth noticing how the paradoxical sting of the Surprise Examination can be removed: Instead of lamenting the epistemic situation of ideal believers in this particular set-up, we might rather conclude that the express wording of the teacher’s announcement — if taken literally — frustrates his attempt to inform his class partially of an impending (partial surprise) exam.

*Synchronic Case:* Consider first the synchronic case of the Surprise Examination Paradox. On Monday morning, a teacher announces to his class that an exam will be written on Monday ( $E_1$ ) but that it will take the class by surprise ( $\neg B_1(E_1)$ ). As a consequence of AE-Transparency, the class cannot believe both announcements if they are ideal believers ( $(E_1 \wedge \neg B_1(E_1)) \notin K_1$ ). Obviously, the exam could still take the class by surprise if they rejected both announcements or just  $E_1$ . In this sense, the teacher’s announcements might still come true although they cannot both be believed by ideal reasoners. For the teacher, this means that he cannot both announce the exam and call it a surprise. Announcing the date of a surprise exam is not coherent.

*Diachronic Case:* The same situation arises in the diachronic case if the class are also diachronically ideal believers and fulfil  $\text{Iteration}_K$ . On Monday morning, the teacher now announces to his class that an exam will be written either on Monday ( $E_1$ ) or on Tuesday ( $E_2$ ) and that, in either case, it will take the class by surprise. If we also assume that the class believes that they will remember correctly on Tuesday

whether there was no exam on Monday, we arrive at the following list of announcements:

$$\begin{array}{ll} E_1 \vee E_2 & E_1 \rightarrow \neg B_1(E_1) \\ \neg E_1 \rightarrow B_2(\neg E_1) & E_2 \rightarrow \neg B_2(E_2) \end{array} \quad (6)$$

*Observation 5.13.* In the presence of AE–Transparency and Iteration, the announcements (6) cannot all be believed at 1.

As in the synchronic case, the announcements (6) might still come true if the class reject at least one of them. If the teacher intended to communicate to his class (make them believe) that the exam will be written either on Monday or Tuesday and nothing more, he frustrated his own attempt by incorrectly spelling out under what conditions the exam would be a surprise after this announcement. What he probably meant to announce was merely that the exam will be written on Monday or Tuesday. He may then coherently add that it will take the class by surprise, unless it will be written on Tuesday. If the teacher decides by himself that the exam thus announced will be written on Monday, the class cannot predict it reliably. If he decides to set it on Tuesday, the class can predict it on Tuesday but not on Monday. Once an exam is announced (with whatever degree of imprecision), it can no longer come as a complete surprise. If the teacher wants the exam to be a complete surprise no matter when he chooses to set it, why does he tell his class about it in the first place?

#### AUTO–EPISTEMOLOGY AND UPDATING

In analogy to the probabilistic model, we assume that the update rule leads to belief sets in which it is known on what evidence they are based:

$$\text{(EV–Transparency)} \quad \text{Ev}_{i+1}^{\text{tot}}(e) \in K_i^{*i+1}e$$

We say that an auto–epistemic update model is *evidence driven* if and only if changes in beliefs exclusively result from updates with incoming evidence (i.e.,  $K_{\omega, i+1} = K_{\omega, i}^{*i+1} \pi(\omega, i+1)$ ). In evidence driven models,  $\text{Reflection}_K$  takes the form of the following principle:

$$\text{(AE–Revision)} \quad K_i^{*i+1}e = K_i | \text{Ev}_{i+1}^{\text{tot}}(e)$$

*Theorem 5.14.* In evidence driven auto–epistemic update models,  $\text{Reflection}_K$  is equivalent to AE–Revision.

Under Correlation 2, AE-Revision corresponds to AE-Conditionalization. AE-Revision demands that an ideal auto-epistemic reasoner must be informed about his own update methods, whatever they are, but it does not commit the reasoner to any specific methodology. In exactly the same way in which AE-Conditionalization has to be distinguished from Conditionalization, AE-Revision differs from AGM Updating. This identification holds only under an extra postulate which equates updated belief with conditional prior beliefs. For  $M \subseteq \mathcal{L}^{AE}$ , define the operator  $(M)^0 := M \cap \mathcal{L}^0$  that purges sets of sentences from auto-epistemic vocabulary. In the presence of AE-Revision, AGM-Updating is then identical to the following crucial assumption:<sup>27</sup>

$$(\text{CruX}_K) \quad (K_i | \text{Ev}_j^{\text{tot}}(A))^0 = (K_i | A)^0$$

#### FAILURES OF AGM UPDATING

AE-Revision furnishes a manual for translating the properties of an adequate update operator into an auto-epistemic assessment of the given learning situation. On the basis of this translation manual, I will in the remainder of this section try to evaluate the claim that AGM-Updating codifies the ‘logical’ properties of update methods. I will highlight the regimentations that AGM-Updating imposes on the learning situations to which it can be applied adequately.

**Success:** The success condition for updates ( $K^*2$ ) corresponds to subjective reliability of evidence:

$$(\text{Reliability}_K) \quad A \in K_i | \text{Ev}_j^{\text{tot}}(A)$$

If we assume, for simplicity’s sake, that  $A$  and  $\text{Ev}_j^{\text{tot}}(A)$  are compatible with  $K_i$ , failure of  $\text{Reliability}_K$  means according to Observation 5.8 that the reasoner does not believe ‘If I learn  $A$ , then  $A$  is the case’:

$$(\text{Ev}_j^{\text{tot}}(A) \rightarrow A) \notin K_i \quad (7)$$

Under AE-Revision, this non-belief alone is sufficient to overthrow the success condition ( $K^*2$ ), even without a belief  $(\text{Ev}_j^{\text{tot}}(A) \rightarrow \neg A)$  in the deceptiveness of evidence.

**Inclusion:** Inclusion ( $K^*3$ ) corresponds to the following subjective assessment:

$$(K_i | \text{Ev}_j^{\text{tot}}(A))^0 \subseteq (K_i^+ A)^0 \quad (8)$$

If we assume that  $\neg \text{Ev}_j^{\text{tot}}(A) \notin K_i$ , this means for  $B \in \mathcal{L}^0$ :

$$\text{If } (\text{Ev}_j^{\text{tot}}(A) \rightarrow B) \in K_i, \text{ then } (A \rightarrow B) \in K_i. \quad (9)$$

Here is a counter-example against Inclusion:

**(Example 5.4)** Imagine the test of a new cloaking device for an aircraft. You know that in the test the aircraft is flying through a zone that is monitored by radar equipment ( $A \in K_i$ ). The reasoner suspends judgment about the functionality ( $B$ ) of the device ( $B, \neg B \notin K_i$ ). Hence,  $(A \rightarrow \neg B) \notin K_i$ . If the device is functional, then the aircraft will not appear on the radar screen, or so he believes ( $(B \rightarrow \neg \text{Ev}_j^{\text{tot}}(A)) \in K_i$ ). Hence, you violate (9) and Inclusion. (Caution:  $B$  in (9) has been substituted by  $\neg B$ .)

**Relevance:** In the presence of Reduction ( $K_i | \top = K_i$ ), Rational Monotony ( $K^*7$ ) implies Expansion 2 ( $K^*4$ ) which, in turn, implies the Limited Success of updating:

(Limited Success)     If  $\neg A \notin K_i$ , then  $A \in K_i^* A$ .

Rejection of Limited Success must therefore be accompanied by a modification of both Expansion 2 ( $K^*4$ ) and Rational Monotony ( $K^*7$ ). Fuhrmann's (1997) Merge operator attempts to overcome Success ( $K^*2$ ). Fuhrmann rejects ( $K^*7$ ) and replaces ( $K^*4$ ) by the weaker condition of Relevance:<sup>28</sup>

(Relevance)             If  $B \in K_i - K_i^* A$ , then there is a  $\Gamma$  such that

- (i)      $(K_i^* A \cap K_i) \subseteq \Gamma \subseteq K$ ,
- (ii)     $\neg A \notin \text{Cn}(\Gamma)$ , and
- (iii)    $\neg A \in \text{Cn}(\Gamma \cup \{B\})$ .

Relevance is weak enough so as not to imply Limited Success. It requires that, in an update of  $K_i$  with  $A$ ,  $B$  is only given up if — together with the beliefs remaining in the updated belief set —  $B$  implies  $\neg A$ . In other words, hypotheses can only be disconfirmed indirectly through the confirmation of logically conflicting rival hypotheses. Direct disconfirmation is excluded. In particular, Relevance implies that updates with compatible input do not subtract prior beliefs:

$$\text{If } \neg A \notin K_i, \text{ then } K_i \subseteq K_i^* A. \quad (10)$$

Subjectively speaking, Relevance therefore presupposes the following assessment of the learning situation:

$$\text{If } \neg A \notin K_i, \text{ then } K_i \subseteq K_i | \text{Ev}_j^{\text{tot}}(A). \quad (11)$$

The next examples illustrates how (11) and *a fortiori* Relevance can fail:

**(Example 5.5)** Consider again the test of the aircraft cloaking device.

The reasoner believes that the aircraft is flying through the monitored zone ( $A \in K_i$  and hence  $\neg A \notin K_i$ ). On basis of earlier tests he already believes that the device is functional ( $B \in K_i$ ). By  $\text{Iteration}_K$ , he must hence exclude the possibility that the aircraft will appear on the screen ( $\neg \text{Ev}_j^{\text{tot}}(A) \in K_i$ ). But if it is nonetheless detected by the radar system, he has two options: Firstly, he can suspend judgment about the functionality of the device ( $B \notin K_i | \text{Ev}_j^{\text{tot}}(A)$ ), e.g., because he is no longer certain that the test is set up correctly (the radar system could have picked up a civilian aircraft). Secondly, he can conclude that the test is set up correctly and that the device it is malfunctioning ( $\neg B \in K_i | \text{Ev}_j^{\text{tot}}(A)$ ). Both conclusions would violate (11) and Relevance.

**Logical Core:** If the previous counter-examples are efficient, the only surviving axioms are Closure ( $K^*1$ ), Consistency ( $K^*5$ ), Extensionality ( $K^*6$ ), and  $\text{Reduction}_K$ . The above examples demonstrate the extent to which update methods can vary across different learning situations. We thus have to redress the AGM conception of the ‘logical’ properties of update rules. The AGM axioms only characterize a notion of conditional full belief, not of updated full belief. The notion of conditional full belief remains, of course, to be given a suitable interpretation. I have only introduced it as the formal correlate of probabilistic conditional belief. Conditional full beliefs might plausibly be interpreted as the result of hypothetical theorizing or supposing. This would be in line with Skyrms’s (1987b) view that the conditional probability measure  $P_i(\cdot | B)$  specifies the probabilities at  $i$  under the supposition  $B$ .

## Appendix

### Proofs

*Theorem 2.1*

*Ad (i):* For  $i = j$ , Reflection gives us  $P_i(\text{SP}_i(P_i)) = P_i(\text{SP}_i(P_i) | \text{SP}_i(P_i)) = 1$ .

*Ad (ii):* Reflection gives us  $P_k(\text{SP}_k(Q_k) | \text{SP}_i(Q_i)) = P_i(\text{SP}_k(Q_k))$  ( $k \leq i$ ). By AE-Transparency,  $P_k(\text{SP}_k(Q_k)) = 1$  iff  $P_k = Q_k$ . Hence  $P_i(\text{SP}_k(Q_k)) = 1$  iff  $P_k = Q_k$ .  $\square$

*Theorem 2.2*

*Ad (i):* The Total Probability Theorem and NZ-Reflection imply:

$$P_i(\cdot) = \sum_{Q_j \in \mathcal{P}(i,j)} P_i(\text{SP}_j(Q_j)) \times P_i(\cdot | \text{SP}_j(Q_j))$$

$$= \sum_{Q_j \in \mathcal{P}(i,j)} P_i(\text{SP}_j(Q_j)) \times Q_j(\cdot)$$

This implies Iteration:

$$\begin{aligned} E_i(X) &= \sum_{\omega} P_i(\omega) \times X(\omega) \\ &= \sum_{\omega} \left( \left[ \sum_{Q_j} P_i(\text{SP}_j(Q_j)) \times Q_j(\omega) \right] \times X(\omega) \right) \\ &= \sum_{\omega} \sum_{Q_j} P_i(\text{SP}_j(Q_j)) \times Q_j(\omega) \times X(\omega) \\ &= \sum_{Q_j} \sum_{\omega} P_i(\text{SP}_j(Q_j)) \times Q_j(\omega) \times X(\omega) \\ &= \sum_{Q_j} P_i(\text{SP}_j(Q_j)) \times \left( \sum_{\omega} Q_j(\omega) \times X(\omega) \right) \end{aligned}$$

*Ad (ii):* Iteration and AE-Transparency imply:

$$\begin{aligned} P_i(A \cap \text{SP}_j(Q_j)) &= \sum_{Q'_j \in \mathcal{P}(i,j)} P_i(\text{SP}_j(Q'_j)) \times Q'_j(A \cap \text{SP}_j(Q_j)) \\ &= P_i(\text{SP}_j(Q_j)) \times Q_j(A) \end{aligned}$$

NZ-Reflection follows when  $P_i(\text{SP}_j(Q_j)) > 0$ .

*Ad (iii):* Suppose NZ-Reflection and, hence, Iteration hold. Then  $P_i(\cdot)$  is a convex combination of the elements in  $\mathcal{P}(i, j)$ . This implies Generalized NZ-Reflection.

To see that the converse does not hold, consider the following probabilities:

	$A$	$-A$	$\text{SP}_j(P_j^1)$	$\text{SP}_j(P_j^2)$	$\text{SP}_i(P_i)$
$P_i(\cdot)$	.5	.5	.5	.5	1
$P_j^1(\cdot)$	.25	.75	1	0	1
$P_j^2(\cdot)$	.1	.9	0	1	1

When  $P_i(A \cap \text{SP}_j(P_j^1)) \neq \frac{1}{8}$ ,  $P_i(\cdot)$  lies inside the convex closure of  $\mathcal{P}(i, j)$  but violates NZ-Reflection and Iteration.

*Ad (iv):* Simply because probabilities are expectations of the characteristic function, i.e.  $P_i(A) = E_i(\chi(A))$ .<sup>29</sup> To see that the converse does not hold, consider:

	$A$	$B$	$C$	$E(\cdot)$
$P_i(\cdot)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0
$P_j^1(\cdot)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$	-1
$P_j^2(\cdot)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$	-1
$X$	0	-6	6	

These probabilities satisfy Generalized NZ-Reflection but  $E_i(X) = 0$ ,  $E_j^1(X) = -1$ , and  $E_j^2(X) = -1$  violate Generalized Iteration.

*Ad (v):* Define a random variable  $\beta$  that has the form of a fair bet on  $A$  conditional on  $\text{SP}_j(Q_j)$ :

$$\beta = \left\{ \begin{array}{ll} 1 - Q_j(A), & \text{if } A \\ -Q_j(A), & \text{if } -A \end{array} \right\} \Bigg| \text{SP}_j(Q_j)$$

Hence,

$$\begin{aligned} E_j(\beta) &= 0 + P_j(A \cap \text{SP}_j(Q_j)) \times (1 - Q_j(A)) \\ &\quad + P_j(-A \cap \text{SP}_j(Q_j)) \times (-Q_j(A)) \end{aligned}$$

Then there are exactly two cases:

*Case 1:*  $P_j = Q_j$ . AE-Transparency implies  $P_j(A \cap \text{SP}_j(Q_j)) = Q_j(A)$ . Thus,  $E_j(\beta) = Q_j(A)(1 - Q_j(A)) + (1 - Q_j(A))(-Q_j(A)) = 0$ .

*Case 2:*  $P_j \neq Q_j$ . Similarly.

If  $P_i(\text{SP}_j(Q_j)) > 0$ , Generalized Iteration implies  $E_i(\beta) = 0$  and, hence, NZ-Reflection. The structure of this proof is essentially identical to that of the Dutch Book Theorem for NZ-Reflection (cf. van Fraassen (1995), Hild (1997)).

*Ad (vi):* According to Iteration,  $E_{P_i}$  is a convex combination of the elements of the expectations  $E_{Q_j}$  with  $Q_j \in \mathcal{P}(i, j)$ . This implies Generalized Iteration.  $\square$

### Theorem 2.3

In evidence driven models, we have  $\text{SP}_{i+1}(P_{i+1}) = \text{SP}_{i+1}(P_i^{*i+1}e) = \text{SP}_i(P_i) \cap \text{Ev}_{i+1}^{\text{tot}}(e)$ . Reflection can thus simply be re-written as (\*)  $P_i(\cdot | \text{SP}_{i+1}(P_i^{*i+1}e)) = P_i(\cdot)^{*i+1}e$ . On the other hand,  $P_i(\text{SP}_i(P_i)) = 1$  (by AE-Transparency) and hence  $P_i(\text{SP}_{i+1}(P_i^{*i+1}e)) = P_i(\text{Ev}_{i+1}^{\text{tot}}(e))$ . (\*) is therefore equivalent to AE-Conditionalization.  $\square$

### Theorem 3.4

In

Success and Rigidity, replace  $[P_i^{*i+1}e_{GC}](\cdot)$  by  $P_i(\cdot | \text{Ev}_{i+1}^{\text{tot}}(e_{GC}))$ . Note that (by AE-Conditionalization)  $P_i(A|B)^{*i+1}e_{GC} = \frac{P_i(A \cap B)^{*i+1}e_{GC}}{P_i(B)^{*i+1}e_{GC}} = \frac{P_i(A \cap B | \text{Ev}_{i+1}^{\text{tot}}(e_{GC}))}{P_i(B | \text{Ev}_{i+1}^{\text{tot}}(e_{GC}))} = P_i(A|B \cap \text{Ev}_{i+1}^{\text{tot}}(e_{GC}))$ .  $\square$

### Observation 5.8

Notice that  $\text{Cn}(\cdot)$  has the Deduction Property: If  $B \in \text{Cn}(\Gamma \cup \{A\})$ , then  $(A \rightarrow B) \in \text{Cn}(\Gamma)$ . This follows from Logic 2. Now suppose that

$B \in \text{Cn}(M \cup \{A\})$ . Since  $\text{Cn}(\cdot)$  has the Deduction Property,  $(A \rightarrow B) \in \text{Cn}(M)$ . Converse trivial.  $\square$

*Theorem 3.5*

Reliability implies  $P_i(-B_{k'} \cap \text{Ev}_{i+1}^{\text{tot}}(B_{k'})) = 0$  for all  $k' \in K$  for all  $k \neq k'$ , Partitioning (i) and (ii) imply (\*)  $P_i(B_k \cap \text{Ev}_{i+1}^{\text{tot}}(B_{k'})) = 0$  and hence  $P_i(B_k \cap -\text{Ev}_{i+1}^{\text{tot}}(B_k) \cap \text{Ev}_{i+1}^{\text{tot}}(B_{k'})) = 0$ . Since  $\{\text{Ev}_{i+1}^{\text{tot}}(B_k) | k \in K\}$  partitions the support of  $P_i(\cdot)$ , we obtain (\*\*)  $P_i(B_k \cap -\text{Ev}_{i+1}^{\text{tot}}(B_k)) = 0$ . Because of  $\{B_k | k \in K\} = \mathcal{EV}(i, i+1)$ , we have  $P_i(\text{Ev}_{i+1}^{\text{tot}}(B_k)) > 0$  (for all  $k \in K$ ). Together with (\*) and (\*\*), this implies Evidential Independence.  $\square$

*Observation 5.9*

First, suppose that  $(A \wedge B) \in K_i | C$ . Hence,  $B \in K_i | C$  and  $(B \rightarrow A) \in K_i | C$  ( $K|1$ ). The second line of Multiplication implies that  $(A \wedge B) \in K_i | (B \wedge C)$  or  $\neg B \in K_i | (B \wedge C)$ . In the second case,  $K_i | (B \wedge C) = \mathcal{L}$ .

Second, suppose that  $B \in K_i | C$  and  $A \in K_i | (B \wedge C)$ . The second line of Multiplication implies that  $(B \rightarrow A) \in K_i | C$ . Hence,  $(A \wedge B) \in K_i | C$  ( $K|1$ ).  $\square$

*Theorem 5.10*

*Ad (i):* Multiplication and Reduction imply ( $K|3$ ), ( $K|4$ ), ( $K|7$ ), and ( $K|8$ ):

( $K|3$ ) Special case of ( $K|7$ ) under Reduction and ( $K|6$ ).

( $K|4$ ) Special case of ( $K|8$ ) under Reduction and ( $K|6$ ).

( $K|7$ ) Assume that  $C \in K_i | (A \wedge B)$ . Since  $C \models (B \rightarrow C)$ ,  $(B \rightarrow C) \in K_i | (A \wedge B)$  ( $K|1$ ). Multiplication (ii) implies that  $(B \rightarrow C) \in K_i | A$ . It then follows from Observation 5.8 and ( $K|1$ ) that  $K_i | (A \wedge B) \subseteq (K_i | A)^+ B$ .

( $K|8$ ) Assume that  $\neg B \notin K_i | A$  and  $(K_i | A)^+ B$ . Observation 5.8 and ( $K|1$ ) imply that  $(B \rightarrow C) \in K_i | A$ . Multiplication (ii) then implies that  $C \in K_i | (A \wedge B)$ .

*Ad (ii):* Obviously, ( $K|3$ ) and ( $K|4$ ) imply Reduction. The following shows that ( $K|7$ ) and ( $K|8$ ) imply Multiplication: First, suppose that  $(B \rightarrow A) \notin K_i | C$ . Clearly,  $\neg B \notin K_i | C$  ( $K|1$ ). On the other hand, Observation 5.8 and ( $K|1$ ) imply  $A \notin (K_i | C)^+ B$ . It then follows from ( $K|7$ ) that  $A \notin K_i | (B \wedge C)$ .

Second, suppose that  $(B \rightarrow A) \in K_i | C$  and  $\neg B \notin K_i | C$ . From Observation 5.8, ( $K|1$ ), and ( $K|8$ ), we then get  $A \in K_i | (B \wedge C)$ .  $\square$

*Theorem 5.11*

*Ad (i):* The definition of  $\mathcal{K}(i, j)$  together with (K|3), (K|4) yields:  $\forall H_j \in \mathcal{K}(i, j) : K_i | B_j(H_j) = K_i^+ B_j(H_j)$ . Expansive Reflection $_K$  furthermore implies that  $\forall H_j \in \mathcal{K}(i, j) : K_i^+ B_j(H_j) = H_j$ . If therefore  $A \in K_i$ , then  $\forall H_j \in \mathcal{K}(i, j) : A \in H_j$ .

Suppose, on the other hand, that  $\forall H_j \in \mathcal{K}(i, j) : A \in H_j$ . From Expansive Reflection $_K$  and Observation 5.8, we have:  $\forall H_j \in \mathcal{K}(i, j) : (B_j(H_j) \rightarrow A) \in K_i$ . Hence,  $A \in K_i$ . (N.b.: Cn(.) is not compact.)

*Ad (ii):* Assume that  $A \in H'_j$ . Consider the following two cases: First,  $A \in H_j$ . Then  $(B_j(H'_j) \rightarrow A) \in H_j$ . Second,  $A \notin H_j$ . In this case, AE-Transparency implies that  $\neg B_j(H'_j) \in H_j$  and hence  $(B_j(H'_j) \rightarrow A) \in H_j$ . Due to Iteration $_K$ ,  $(B_j(H'_j) \rightarrow A) \in K_i$ .

Now assume that  $(B_j(H'_j) \rightarrow A) \in K_i$ . Iteration $_K$  entails that  $\forall H_j \in \mathcal{K}(i, j) : (B_j(H'_j) \rightarrow A) \in H_j$ . Hence,  $A \in H'_j$ . (N.b.: Cn(.) is not compact.)

If  $H'_j \in \mathcal{K}(i, j)$ , Observation 5.8 then implies Expansive Reflection $_K$ .  $\square$

*Observation 5.12*

Obvious for Generalized Reflection $_K$ , Expansive Reflection $_K$ , and Reflection $_K$ . Consider Iteration and suppose that  $P_i(A) < 1$ . Then there is a  $Q_j \in \mathcal{P}(i, j)$  such that  $Q_j(A) < 1$ . If, on the other hand,  $P_i(A) = 1$ , then for all  $Q_j \in \mathcal{P}(i, j) : Q_j(A) = 1$ . Hence,  $A \in K_{P_i(\cdot)}$  if and only if for all  $A \in K_{Q_j}$  for all  $Q_j \in \mathcal{P}(i, j)$ .  $\square$

*Observation 5.13*

The class cannot exclude the possibility that they will believe  $E_1$  on Tuesday, i.e.  $\neg B_2(E_1) \notin K_1$ . Otherwise,  $E_1 \in K_1$  (Iteration) and we would arrive at a contradiction since also  $B_1(E_1) \in K_1$  (AE-Transparency $_K$ ). Hence, there is a  $H_2 \in \mathcal{K}(1, 2)$  such that  $\neg E_1 \in H_2$ . Suppose that (6) are in  $K_1$ . Iteration implies that the conjunction of (6) is in  $H_2$ . For this reason,  $E_2 \in H_2$  and  $B_2(E_2) \notin H_2$ . But this would violate the AE-Transparency $_K$  of  $H_2$ . Hence, (6) cannot all be in  $K_1$ .  $\square$

*Theorem 5.14*

In evidence driven models, we have  $B_{i+1}(K_{i+1}) \Leftrightarrow B_{i+1}(K_i^{*i+1}e) \Leftrightarrow B_i(K_i) \wedge \text{Ev}_{i+1}^{\text{tot}}(e)$ . Reflection $_K$  can thus simply be re-written as (\*)  $K_i | B_{i+1}(K_i^{*i+1}e) = K_i^{*i+1}e$ . On the other hand,  $B_i(K_i) \in K_i$  (by AE-Transparency $_K$ ) and hence  $(B_{i+1}(K_i^{*i+1}e) \leftrightarrow \text{Ev}_{i+1}^{\text{tot}}(e)) \in K_i$ . (\*) is therefore equivalent to AE-Revision.  $\square$

## Notes

<sup>1</sup> The following inference is a standard example of non-monotonic default reasoning: If I believe that Tweety is a bird, I conclude — in the absence of a belief to the contrary (and by default) — that Tweety can fly. But should my knowledge increase and imply that Tweety is a penguin, I will retract this conclusion (cf. Moore (1985)). Konolige (1987) shows how to reduce non-monotonic default logic to auto-epistemic logic.

<sup>2</sup> Van Fraassen (1995) claims that Conditionalization implies Generalized Reflection and that Generalized Reflection implies Reflection.

<sup>3</sup> The index  $I$  roughly resembles a time index, although the location in time of an epistemic state is not of primary interest. A certain time might elapse between two immediately subsequent epistemic states.

<sup>4</sup> Conditional probability functions that satisfy (CP1)–(CP4) are called ‘full’. They can be represented by ‘dimensionally well-ordered’ families of unconditional probability functions. Cf. Spohn (1986) and van Fraassen (1976). Most mathematical textbooks define conditional probabilities merely as a shorthand for the ratio (1) of unconditional probabilities.

<sup>5</sup> These are Shafer’s situations trees; see Shafer (1985), (1996).

<sup>6</sup> For more details cf. Section 5.

<sup>7</sup> If we simply wrote Reflection as ‘ $P_i(A|SP_j(Q)) = Q(A)$ ’, (where  $Q \in \mathcal{PR}$  and  $i \leq j$ ) it would follow that  $Q(SP_j(Q)) = P_i(SP_j(Q)|SP_j(Q)) = 1$  where  $Q$  can be a personal probability function at any arbitrary stage; especially  $P_i(SP_j(P_i)) = 1$  for  $i \neq j$ . Similarly, if Reflection were not restricted to trajectories one of which branches off the other).

<sup>8</sup> Roughly speaking, the expectation operator can be diachronically iterated: ‘ $E_i(X) = E_i(E_j(X))$ ’.

<sup>9</sup> A version of Generalized Reflection has also been considered by Skyrms (1987a).

<sup>10</sup> If we also consider probability measures with non-countable support, some complications arise from the fact that the conditional probabilities used in the formulation of Reflection cannot be defined as ratios of unconditional probabilities when the condition has zero measure. Generalized Iteration and Generalized Reflection carry over completely unmodified. In the formulation of Iteration, we simply need to replace the sums by Lebesgue integrals.

<sup>11</sup> Two pieces of evidence  $e$  and  $e'$  are update-equivalent if and only if  $Q^*e = Q^*e'$  for all  $Q \in \mathcal{PR}$ . We assume that  $\mathcal{EV}$  identifies update-equivalent pieces of evidence with each other. That can always be achieved by replacing pieces of evidence with their equivalence class under update-equivalence.

<sup>12</sup> The restriction to  $\mathcal{A}^0$  is necessary since (Generalized) Conditionalization does not automatically ensure Transparency.

<sup>13</sup> For a very similar theorem cf. Skyrms (1980). Skyrms (1987b) provides conditions under which AE-Conditionalization validates Entropy Maximization.

<sup>14</sup> The support of  $P_i(\cdot)$  is the set of all possible worlds with non-zero probability. This present partition  $\{B_k|k \in K\}$  must not be confused with the partition in a Jeffrey constraint.

<sup>15</sup> Generalized Conditionalization is *not* a special case of Conditionalization. I argue that total evidence and, hence, update rules are relative to some given protocol (evidence function). Generalized Conditionalization and simple Conditionalization deal with different types of evidence and different protocols. Because of limitations in the expressiveness of  $\mathcal{A}^0$ , it may not always be possible to transform evidence in the form of a Jeffrey constraint into a single proposition  $B \in \mathcal{A}^0$ . Hence, Generalized Conditionalization does not reduce to simple Conditionalization.

<sup>16</sup> An exposition of related puzzles can be found in M.Bar-Hillel/Falk (1982)

<sup>17</sup> In the Three Prisoners' Puzzle, Prisoner 1 receives evidence in form of an utterance made to her by a guard ( $\pi$ ). The guard tells Prisoner 1 that Prisoner 2 is going to be shot ( $E$ ). Jeffrey has Prisoner 1 conditionalize on the proposition 'Prisoner 1 is told that Prisoner 2 is going to be shot' ( $E^*$ ), rather than on the proposition 'Prisoner 2 is going to be shot' ( $E$ ). This means that Jeffrey changes the underlying protocol and thereby trivially obtains Conditionalization. In my view, the Three Prisoners' Puzzle is a counter-example to Conditionalization because we have to stick to the initially chose protocol  $\pi$ . Even after Jeffrey's switch to the new protocol  $\pi^*$ , it is again possible to construct a Three Prisoners' Puzzle\* that violates Conditionalization relative to  $\pi^*$ .

<sup>18</sup> This uniform distribution can either be built up from scratch assuming that the remaining five possibilities are equiprobable or it can result from the equiprobable distribution over the original six possible states plus Conditionalization on the new information that at least one ace has been drawn.

<sup>19</sup>  $\mathcal{L}$  is assumed to be closed under truth-functional connectives. ' $A \iff B$ ' abbreviates ' $\text{Cn}(A) = \text{Cn}(B)$ '.  $\top$  and  $\perp$  are introduced as abbreviations for arbitrary tautologies and contradictions ( $\top \in \text{Cn}(\emptyset)$ ,  $\text{Cn}(\perp) = \mathcal{L}$ ).

<sup>20</sup> The second condition also implies that  $\text{Cn}(\cdot)$  has the Deduction Property: If  $B \in \text{Cn}(\Gamma \cup \{A\})$ , then  $(A \rightarrow B) \in \text{Cn}(\Gamma)$ . When  $\mathcal{A}$  and  $\mathcal{L}$  are sufficiently expressive,  $\text{Cn}(\cdot)$  may only be able to satisfy the second condition if it is not compact.

<sup>21</sup> To obtain a propositional belief sets, define  $\mathbf{K}_Q := \bigcap \{ \mathbf{A} \in \mathcal{A} \mid Q(\mathbf{A}) = 1 \}$ .

<sup>22</sup> Because the 'absurd belief set'  $\mathcal{L}$  is not a belief set according to our definition, some minor adjustments of the AGM axioms are in place. I write ' $K_i|A$ ' rather than ' $K_i^*A$ ' in order to avoid the pre-emptive identification of revisions with update rules.

<sup>23</sup> If probabilities measures range over sentences rather than events, the axioms for conditional probabilities must be supplemented by corresponding extensionality axiom:  $P_i(\cdot|A) = P_i(\cdot|B)$ , if  $A \iff B$ .

<sup>24</sup> Multiplication $_K$ , Reduction $_K$ , and  $(K|6)$  imply the translation of (CP5):

$$\begin{aligned} B \in K_i \text{ and } A \in K_i|B &\text{ iff } (A \wedge B) \in K_i \\ \neg B \notin K_i \text{ and } A \notin K_i|B &\text{ iff } (B \rightarrow A) \notin K_i \end{aligned}$$

<sup>25</sup> AE-Transparency $_K$  corresponds to Moore's (1985) definition of 'stable sets'.

<sup>26</sup> Iteration implies: If  $A \in K_i$ , then  $B_j(A) \in K_i$ , for  $i \leq j$ . Phrased in terms of a modal logic of belief, this corresponds to Binkley's (1968) principle ' $B_i(A) \rightarrow B_i B_j(A)$ '.

<sup>27</sup> Again, the restriction to  $\mathcal{L}^0$  is necessary because AGM updates do not automatically take care of auto-epistemic principles like Transparency.

<sup>28</sup> Fuhrmann confines himself to the 'basic axioms' without  $(K^*7)$  and  $(K^*8)$ . Nor does he presuppose Closure  $(K^*1)$ .

<sup>29</sup> The characteristic function of a sentence  $A$  is defined as

$$\chi = \begin{cases} 1, & \text{if } A \\ 0, & \text{otherwise.} \end{cases}$$

## References

- Bar-Hillel, M., Falk, R. (1982), "Some Teasers Concerning Conditional Probabilities", *Cognition* **11**, 109-122
- Binkley, R. (1968), "The Surprise Examination in Modal Logic", *J Phil* **65**, 127-136
- Diaconis, P., Zabell, S.L. (1982), "Updating Subjective Probability", *Journal of the American Statistical Association* **77**, 822-830

- Freund, J.E. (1965), "Puzzle or Paradox?", *American Statistician* **19**, 29-44. For the discussion that followed cf. **20**, 34-37 and **21**, 38-51
- Fuhrmann, A. (1995), "Revision, Merge, and Inference", *forthcoming*
- Gärdenfors, P. (1988), *Knowledge in Flux*, Cambridge (Mass.)/London: MIT
- Goldstein, M. (1983), "The Prevision of Prevision", *Journal of the American Statistical Association* **78**, 817-819
- Hacking, I. (1967), "Slightly More Realistic Personal Probabilities", *Phil Sci* **34**, 311-325
- Hild, M. (1997), "The Coherence Argument Against Conditionalization", *forthcoming*
- Hintikka (1962), *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, Ithaca (NY): Cornell University
- Jeffrey, R.C. (1983), *The Logic of Decision*, 2nd edn, Chicago: University of Chicago Press
- Jeffrey, R.C. (1988), "Conditioning, Kinematics, and Exchangeability", in: Skyrms, B., Harper, W.L. (eds.), *Causation, Chance, and Credence*, Kluwer, **Vol. 1**, 221-255
- 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
- Maher, P. (1993), *Betting on Theories*, Cambridge: CUP
- Moore, R.C. (1985), "Semantical Considerations on Nonmonotonic Logic", *Artificial Intelligence* **25**, 75-94
- Shafer, G. (1985), "Conditional Probability", *International Statistical Review* **53**, 261-277
- Shafer, G. (1996), "Can the Various Meanings of Probability be Reconciled?", *forthcoming*
- Skyrms, B. (1980), *Causal Necessity*, New Haven/London: Yale University
- Skyrms, B. (1987a), "Dynamic Coherence and Probability Kinematics", *Phil Sci* **54**, 1-20
- Skyrms, B. (1987b), "Updating, Supposing, and Maxent", *Theory and Decision* **22**, 225-246
- Spohn, W. (1978), *Grundlagen der Entscheidungstheorie*, Kronberg (Taunus): Scriptor
- Spohn, W. (1986), "The Representation of Popper Measures", *Topoi* **5**, 69-74
- Teller, P. (1973), "Conditionalization and Observation", *Synthese* **26**, 218-258
- van Fraassen, B.C. (1976), "Representation of Conditional Probabilities", *J Phil Log* **5**, 417-430
- van Fraassen, B.C. (1983), "Shafer on Conditional Probability", *J Phil Log* **12**, 467-470
- 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
- Wright, C., Sudbury, A. (1977), "The Paradox of the Unexpected Examination", *Aust J Phil* **55**, 41-58

*Address for correspondence:* Holywell Manor, Manor Road, Oxford OX1 3UH (GB),  
email: matthias.hild@ox.ac.uk