Bayes rules all: On the equivalence of various forms of learning in a probabilistic setting

Balázs Gyenis
Institute of Philosophy, HAS RCH
30 Orszaghaz str, Budapest, Hungary
✉ gyepi@hps.elte.hu 🌐 hps.elte.hu/~gyepi

December 26, 2014

Abstract

Jeffrey conditioning is said to provide a more general method of assimilating uncertain evidence than Bayesian conditioning. We show that Jeffrey learning is merely a particular type of Bayesian learning if we accept either of the following two observations:

- Learning comprises both probability kinematics and proposition kinematics.
- What can be updated is not the same as what can do the updating; the set of the latter is richer than the set of the former.

We address the problem of commutativity and isolate commutativity from invariance upon conditioning on conjunctions. We also present a disjunctive model of Bayesian learning which suggests that Jeffrey conditioning is better understood as providing a method for incorporating *unspecified* but *certain* evidence rather than providing a method for incorporating *specific* but *uncertain* evidence. The results also generalize over many other subjective probability update rules, such as those proposed by Field (1978) and Gallow (2014).

Keywords: Bayesian learning, Jeffrey conditioning, Gallow conditioning, commutativity, formal epistemology.

1 Introduction

Jeffrey conditioning is a way of obtaining a new probability P_2 from an old probability P_1 and from new probability values q_i ($q_i \geq 0$, $\sum_i q_i = 1$) pre-assigned to a partition $\{E_i\}_i$, $P_1(E_i) > 0$ by making use of the *Jeffrey formula*:

$$P_2(H) = \sum_i q_i \cdot P_1(H|E_i). \tag{1}$$

If we assume that in P_2 an element E of the partition becomes certain (that is, if $P_2(E) = q_E = 1$) the Jeffrey formula (1) reduces to the *Bayes formula*

$$P_2(H) = P_1(H|E), \tag{2}$$

according to which the new probability P_2 is obtained from P_1 by *(Bayesian) conditioning* upon an evidence E whose prior probability is non-zero $P_1(E) > 0$.

On the basis that the Jeffrey formula subsumes the Bayes formula as a special case while it also seems to allow for conditioning upon non-certain evidence it is frequently asserted in the literature that Jeffrey conditioning provides a "more general method of assimilating uncertain evidence" than Bayesian conditioning (Jeffrey; 1983, p. 171). Since the inference ostensibly proceeds from a contrast between mathematical formulas to a contrast between accounts of learning it clearly hinges on assumptions about how the mathematical formulas get incorporated into the respective accounts of learning. We argue that, contrary to the intended conclusion, Jeffrey learning, as well as many other models of learning (including that of Adams, Field (1978), Gallow (2014) etc.), can be properly understood as a particular type of Bayesian learning, and Bayesian conditioning is able to accommodate all these subjective belief revision schemes.

In its most basic form the mathematical structure Bayesianism uses to capture the evidential reasoning of agents is a probability space. A probability space has two components: a space of propositions $\mathcal{L}$ and a probability measure P defined on $\mathcal{L}$. Correspondingly there are two ways in which learning can be mathematically modeled in this basic framework:

(i) *Probability kinematics* tells us how assignments of probability can change.

(ii) *Proposition kinematics* tells us how the space of propositions can change.

To get a Bayesian account of learning one needs to specify how probability kinematics *and* proposition kinematics works. The Bayes and the Jeffrey formulas address probability kinematics by keeping the space of propositions $\mathcal{L}$ fixed. Neither of these formulas address proposition kinematics. In particular, they do not address the possibility of refining the space of propositions. Refining the space of propositions – in other words, extending the probability space (see later) – can be interpreted as the agent learning new distinctions which she was not able to express before. Examples include the introduction of new terms to the language: probabilities of propositions which do not make use of the newly introduced terms remain unchanged, but the agent will now be able to formulate new propositions as well, on which in turn she may also be able to condition.

Section 2 shows that as long as one does not conflate a single component of Bayesian learning, namely probability kinematics, with the whole Bayesian account of incorporating evidence, Jeffrey learning can be properly understood as a particular type of Bayesian learning. Section 3 illustrates the concepts through a simple example and explains why non-commutativity of Jeffrey updates is not a problem after all.

Section 4 tackles probability kinematics itself. A mathematical model of updating subjective beliefs in propositions on the basis of incoming evidence requires two representational elements:

(i) a set of propositions $\mathcal{L}$,

(ii) a set of representations of evidences $\mathcal{S}$ on the basis of which the agent may update her subjective beliefs in propositions $\mathcal{L}$.

Section 4 argues that by adopting seemingly reasonable assumptions regarding $\mathcal{L}$ and $\mathcal{S}$ one can again reach the conclusion that Jeffrey learning can be understood as a particular type of Bayesian learning, even without taking proposition kinematics into account.

Section 5 takes a detour to show that agents who may be called non-maximal Bayesian face a paradox of future dependence of conditioning on conjunctions; along the way we isolate two important desiderata regarding learning models that are often meshed together: commutativity and invariance upon conditioning on conjunctions. Finally Section 6 proposes a disjunctive model of Bayesian learning and comments on the irreversibility of Bayesian conditioning. All proofs are delegated to the Appendix.

2 Probability and proposition kinematics

In what follows P_1 and P_2 are σ-additive probabilities defined on the same Boolean σ-algebra $\mathcal{L}$ with unit element Ω. It is assumed throughout that the probability P_1 is not trivial, i.e. there is at least one element whose probability is neither zero nor unity in P_1.

Definition 2.1 *We say that* P_2 can be obtained from P_1 by Bayesian conditioning without extension *if there exists an* $A \in \mathcal{L}$, $P_1(A) > 0$ *such that for all* $H \in \mathcal{L}$:

$$P_2(H) = P_1(H|A). \tag{3}$$

Definition 2.2 *We say that* P_2 can be obtained from P_1 by (finite) Jeffrey conditioning without extension *if there exists a partition (with finitely many elements)* $\{E_i\}_i$ *of* Ω *with* $P_1(E_i) > 0$ *and* $q_i \geq 0$, $\sum_i q_i = 1$ *such that for all* $H \in \mathcal{L}$:

$$P_2(H) = \sum_i q_i \cdot P_1(H|E_i). \tag{4}$$

As mentioned in the introduction, Jeffrey conditioning without extension is more general than Bayesian conditioning without extension:

Proposition 2.1 *If* P_2 *can be obtained from* P_1 *by Bayesian conditioning without extension, then* P_2 *can be obtained from* P_1 *by finite Jeffrey conditioning without extension.*

Counterexample 2.1 *There is an example where* P_2 *can be obtained from* P_1 *by finite Jeffrey conditioning without extension but where* P_2 *can not be obtained from* P_1 *by Bayesian conditioning without extension.*

We now also take into consideration proposition kinematics and allow for the possibility of refining our space of propositions.

Definition 2.3 *The probability space* $(\tilde{\Omega}, \tilde{\mathcal{L}}, \tilde{P})$ *is an* extension *of the probability space* $(\Omega, \mathcal{L}, P)$ *if there exists a probability preserving injective Boolean algebra homomorphism* $\tilde{}$ *from* $\mathcal{L}$ *to* $\tilde{\mathcal{L}}$.

Definition 2.4 *We say that* P_2 can be obtained from P_1 by Bayesian conditioning with extension *if there exists an extension* $(\tilde{\Omega}, \tilde{\mathcal{L}}, \tilde{P}_1)$ *of* $(\Omega, \mathcal{L}, P_1)$ *and an* $\hat{A} \in \tilde{\mathcal{L}}$, $\tilde{P}_1(\hat{A}) > 0$ *such that for all* $H \in \mathcal{L}$:

$$P_2(H) = \tilde{P}_1(\tilde{H}|\hat{A}). \tag{5}$$

Definition 2.5 *We say that* P_2 can be obtained from P_1 by (finite) Jeffrey conditioning with extension *if there exists an extension* $(\tilde{\Omega}, \tilde{\mathcal{L}}, \tilde{P}_1)$ *of* $(\Omega, \mathcal{L}, P_1)$, *a partition (with finitely many elements)* $\{\hat{E}_i\}_i$ *of* $\tilde{\Omega}$ *with* $\tilde{P}_1(\hat{E}_i) > 0$ *and* $q_i \geq 0$, $\sum_i q_i = 1$ *such that for all* $H \in \mathcal{L}$:

$$P_2(H) = \sum_i q_i \cdot \tilde{P}_1(\tilde{H}|\hat{E}_i). \tag{6}$$

Proposition 2.2 *If P_2 can be obtained from P_1 by finite Jeffrey conditioning without extension then P_2 can be obtained from P_1 by Bayesian conditioning with extension.*

Proposition 2.2 is a consequence of a slight generalization of Theorem 2.1 of the seminal paper of Diaconis and Zabell (1982) whose implications would probably have been better appreciated if their Theorem 2.2 emphasized more clearly the reliance on non-finite partitions (see Counterexample 2.3 and Corollary 4.2 for further details). The converse of Proposition 2.2 does not hold:

Counterexample 2.2 *There is an example where P_2 can be obtained from P_1 by Bayesian conditioning with extension, but P_2 can not be obtained from P_1 by finite Jeffrey conditioning without extension.*

However when extension is allowed on both ends the two notions of conditioning have the same power:

Proposition 2.3 *P_2 can be obtained from P_1 by finite Jeffrey conditioning with extension if and only if P_2 can be obtained from P_1 by Bayesian conditioning with extension.*

It is not just mathematically justified to talk about elements of the refined spaces of Proposition 2.2 and 2.3 as 'propositions'. For instance when $\mathcal{L}$ is generated by taking the smallest Boolean σ-algebra of a set of propositions of a formal language it is clear from the proofs that an $\tilde{\mathcal{L}}$ generated after the addition of a logically independent proposition can form the basis of the required extension.

Proposition 2.2 shows that an agent can always refine her space of propositions in such a way that her Jeffrey conditioning corresponds to a Bayesian conditioning upon a proposition $\hat{A}$ from the refined space. This includes the elements of the partition: $q_i = P_2(E_i) = \tilde{P}_1(\tilde{E}_i|\hat{A})$ for all $i = 1, ..., n$. From this we can see that Bayesian learning can adequately accommodate Jeffrey learning and there is no need for any new form of 'input law' or experience-representing structure such as the one sought by Field (1978). (Field's proposal have been criticized on other grounds, see i.e. Garber (1980).) The new information, the "Bayesian factor" can simply come in the form of a proposition like $\hat{A}$ and the agent learns how to update her other propositions via basic Bayesian conditioning upon such $\hat{A}$.

Analogues of Proposition 2.2 can also be obtained for updating rules other than Jeffrey's, such those of Field (1978) and Gallow (2014). Hence we can also think of these alternative update rules as being adequately accounted for by Bayesian conditioning, albeit we may need to do the conditioning upon

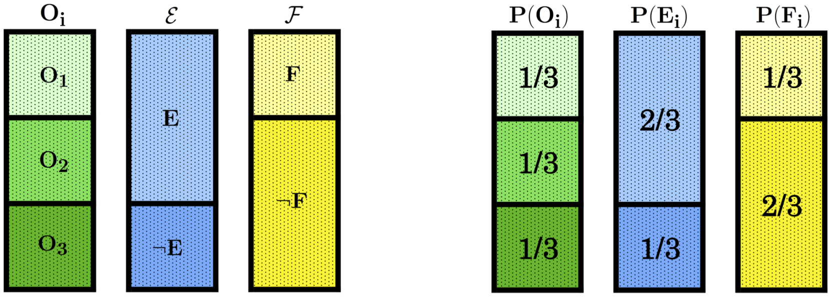

Figure 1: A uniform probability space with three atoms O_1, O_2, O_3 and two element partitions $\mathcal{E} = \{E, \neg E\} = \{\{O_1, O_2\}, O_3\}$, $\mathcal{F} = \{F, \neg F\} = \{O_1, \{O_2, O_3\}\}$.

a proposition from an extended space. Analogues of Proposition 2.3 also immediately follow when the alternative update rule reduces to Bayesian conditioning with a suitable choice of parameters. See Proposition 4.1 and Corollary 4.1 for details.

The condition that the partition is finite in Proposition 2.2 is necessary:

Counterexample 2.3 *There is an example where P_2 can be obtained from P_1 by (non-finite) Jeffrey conditioning without extension, but P_2 can not be obtained from P_1 by Bayesian conditioning with extension.*

The counterexample shows that Jeffrey conditioning can not be obtained in general as Bayesian conditioning if we allow the partition of propositions whose new probability we set by hand to be non-finite. Proposition 4.2 and Corollary 4.2 however shows that in a precise approximate sense Bayesian conditioning is still capable of capturing non-finite Jeffrey conditioning. (It would be possible but cumbersome to formulate an analogue of Corollary 4.2 in the framework of this section and hence omitted.)

3 The problem of commutativity

We illustrate the claims of the previous section with a simple example. Let E be the proposition that *"It is going to rain"* and F be the proposition that *"RoboCop is going to get wet"*, and suppose that both Alice and Bob initially maintains that the chance of raining is 2/3, the chance of RoboCop getting wet given that it rains is 1/2, and the chance of RoboCop getting wet given that it does not rain is 0. As one can quickly check this information can be represented in a uniform probability space with three atoms $O_1 = E \wedge F$, $O_2 = E \wedge \neg F$, and $O_3 = \neg E \wedge \neg F$. (See Figure 1.)

With Jeffrey conditioning *order-dependence* or *non-commutativity* of successive updates is a well known concern (see for instance Döring (1999)). Suppose that Alice first Jeffrey conditions P to $P^{\mathcal{E}}$ with partition $\mathcal{E}$ and corresponding probability values q_i and then she performs on $P^{\mathcal{E}}$ a second Jeffrey conditioning with partition $\mathcal{F}$ and corresponding probability values r_j to get $P^{\mathcal{F}\mathcal{E}}$. Bob performs Jeffrey conditionings on the same partitions and probability values, but does so in the reverse order: first he Jeffrey conditions P to $P^{\mathcal{F}}$ with partition $\mathcal{F}$ and values r_j, and then he Jeffrey conditions $P^{\mathcal{F}}$ to $P^{\mathcal{E}\mathcal{F}}$ with partition $\mathcal{E}$ and values q_i. It turns out that unless $\mathcal{E}$ and $\mathcal{F}$ are so-called Jeffrey independent with respect to P, q_i, and r_j (see Theorem 3.2 of Diaconis and Zabell (1982)) the result depends on the order in which Alice and Bob updates: $P^{\mathcal{F}\mathcal{E}} \neq P^{\mathcal{E}\mathcal{F}}$.

Thus the order in which Jeffrey conditionings happen does in general matter. On the other hand it seems reasonable to maintain that the order in which different pieces of evidence arrive should not matter in the resulting change of subjective beliefs. So non-commutativity is a problem if one wants to think of the partition and the associated new probability values of the Jeffrey formula as direct representations of the evidence that is supposed to be incorporated by the agent.

For a concrete example let's assume that in the first step Alice decreases her belief in the coming rain (E) to 1/2 and in the second step she increases her belief in RoboCop getting wet (F) to 1/2, while in the first step Bob increases his belief in RoboCop getting wet to 1/2 and in the second step he decreases his belief in the coming rain to 1/2. It is easy to check (see Figure 2) that the corresponding Jeffrey updates do not commute, i.e. $P^{\mathcal{F}\mathcal{E}}(O_1) = 1/2$ while $P^{\mathcal{E}\mathcal{F}}(O_1) = 1/3$, and hence Alice and Bob end up with different beliefs.

According to the intended interpretation Alice first learns that a specific proposition, namely that it is going to rain, has a chance 1/2. But how does this happen? Alice may look out in the window and see that the sky is clear; or she may gather this information from the barometer in her room; or she may hear about it from a radio broadcast. From the mere fact that Alice alters her beliefs in the coming rain to 1/2 we do not get a definite answer to the question which among these possible causes prompts Alice to alter her beliefs, nevertheless something does. In the spirit of Proposition 2.2 we can conceive of the actual causal influence on Alice as Alice learning a proposition (potentially from a refined probability space) with certainty.

Suppose now that we know the actual influences: Alice first decreases her belief in the coming rain to 1/2 due to seeing that the sky is relatively clear, and second she increases her belief in RoboCop getting wet to 1/2 due to seeing that RoboCop's umbrella has holes. Meanwhile Bob first increases

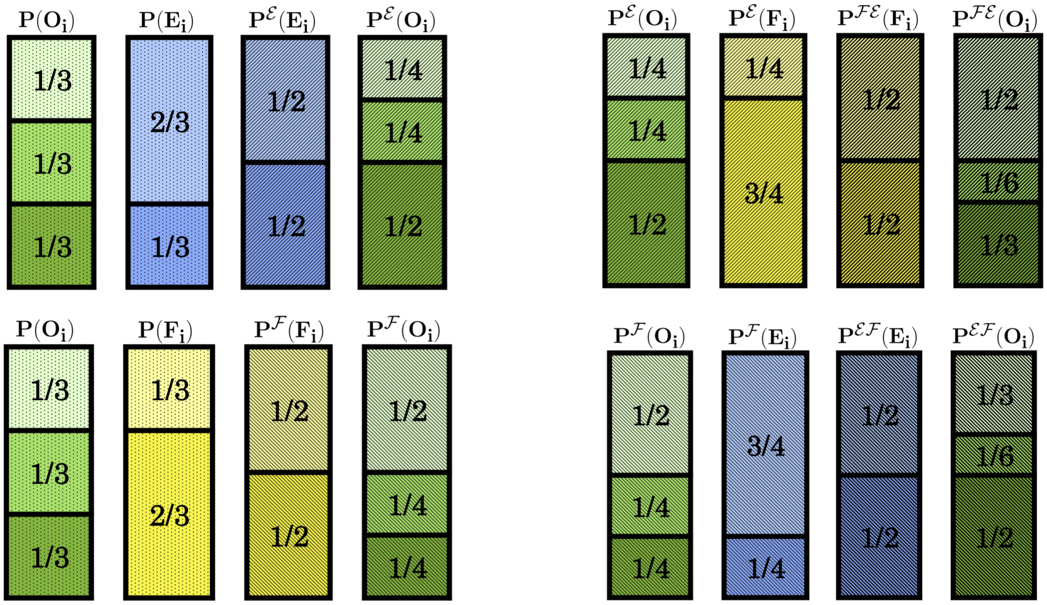

Figure 2: Illustration of successive Jeffrey conditionings. Alice and Bob Jeffrey conditions on the same partitions with the same new probability values but in a different order. $P^{\mathcal{E}}(E) = P^{\mathcal{E}\mathcal{F}}(E) = 1/2$ and $P^{\mathcal{F}}(F) = P^{\mathcal{F}\mathcal{E}}(F) = 1/2$. We see that Jeffrey conditioning is not commutative: $P^{\mathcal{F}\mathcal{E}} \neq P^{\mathcal{E}\mathcal{F}}$, i.e. $P^{\mathcal{F}\mathcal{E}}(O_1) = 1/2$ while $P^{\mathcal{E}\mathcal{F}}(O_1) = 1/3$.

his beliefs in RoboCop getting wet to 1/2 due to seeing that RoboCop's umbrella has holes, and second he decreases his belief in the coming rain to 1/2 due to seeing that the sky is relatively clear. How it is possible that they arrive to different beliefs, even though they saw "the same things"?

We argue that the intuition expressed by Osherson (2002) for why these Jeffrey updates do not (and should not be expected to) commute can be made precise and generalized: Alice and Bob could not have seen the same sky and the same umbrella holes. To illustrate consider an extension our probability space with 14 atoms; for simplicity we omit the tildes from above the extended probabilities. With

$\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$ corresponding to O_1,

$\{\omega_6, \omega_7, \omega_8, \omega_9, \omega_{10}\}$ corresponding to O_2,

$\{\omega_{11}, \omega_{12}, \omega_{13}, \omega_{14}\}$ corresponding to O_3,

and their probabilities given by Figure 3, the smallest probability space containing $\{\omega_i\}_{i=1}^{14}$ is an extension of our original probability space.

As one can easily verify (see Figure 3 and Figure 4 for details) we can obtain the $P^{\mathcal{E}}$, $P^{\mathcal{F}\mathcal{E}}$, $P^{\mathcal{F}}$, and $P^{\mathcal{E}\mathcal{F}}$ Jeffrey updated probabilities of Alice and Bob from P by Bayesian conditioning with this

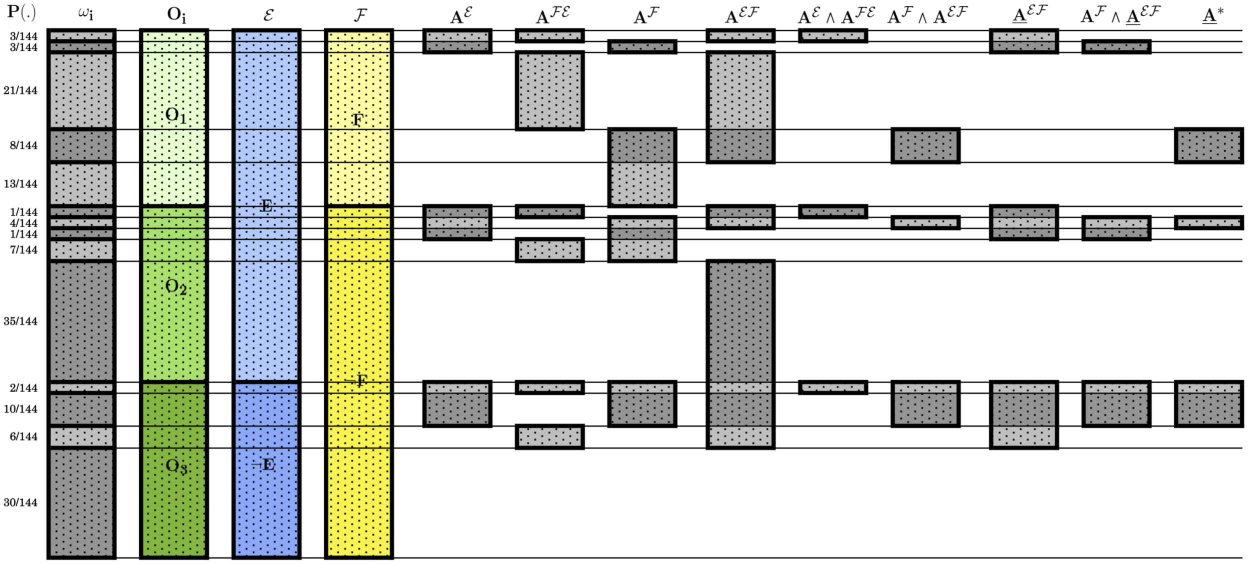
$\mathbf{P}(.)$
$\omega_{\mathbf{i}}$
$\mathbf{O_i}$
$\mathcal{E}$
$\mathcal{F}$
$\mathbf{A}^{\mathcal{E}}$
$\mathbf{A}^{\mathcal{FE}}$
$\mathbf{A}^{\mathcal{F}}$
$\mathbf{A}^{\mathcal{EF}}$
$\mathbf{A}^{\mathcal{E}} \wedge \mathbf{A}^{\mathcal{FE}}$
$\mathbf{A}^{\mathcal{F}} \wedge \mathbf{A}^{\mathcal{EF}}$
$\underline{\mathbf{A}}^{\mathcal{EF}}$
$\mathbf{A}^{\mathcal{F}} \wedge \underline{\mathbf{A}}^{\mathcal{EF}}$
$\underline{\mathbf{A}}^{*}$
3/144
3/144
21/144
8/144
13/144
1/144
4/144
1/144
7/144
35/144
2/144
10/144
6/144
30/144
$\mathbf{O_1}$
$\mathbf{O_2}$
$\mathbf{O_3}$
$\mathbf{E}$
$\neg\mathbf{E}$
$\mathbf{F}$
$\neg\mathbf{F}$

Figure 3: The extended probability space with definitions of various elements.

extension. In other words there exist $A^{\mathcal{E}}, A^{\mathcal{FE}}, A^{\mathcal{F}}, A^{\mathcal{EF}}$ in our extended probability space such that for all H in the original probability space we have[1]

$$\begin{aligned} P^{\mathcal{E}}(H) &= P_{A^{\mathcal{E}}}(H) && (7)\\ P^{\mathcal{FE}}(H) &= P^{\mathcal{E}}_{A^{\mathcal{FE}}}(H) && (8)\\ P^{\mathcal{F}}(H) &= P_{A^{\mathcal{F}}}(H) && (9)\\ P^{\mathcal{EF}}(H) &= P^{\mathcal{F}}_{A^{\mathcal{EF}}}(H). && (10) \end{aligned}$$

Thus $A^{\mathcal{E}}$ is a possible representation of the evidence that Alice learns with certainty in the first step and $A^{\mathcal{FE}}$ is a possible representation of the evidence that Alice learns with certainty in the second step. Similarly $A^{\mathcal{F}}$ is a possible representation of the evidence that Bob learns with certainty in the first step and $A^{\mathcal{EF}}$ is a possible representation of the evidence that Bob learns with certainty in the second step.

The example also shows why the Jeffrey updates fail to commute: the Bayesian factors that induce the change of subjective beliefs are not the same when the update happens in different order, that is $A^{\mathcal{E}} \neq A^{\mathcal{EF}}$ and $A^{\mathcal{F}} \neq A^{\mathcal{FE}}$. (And as one could also check, $P_{A^{\mathcal{E}}} \neq P_{A^{\mathcal{EF}}}$ and $P_{A^{\mathcal{F}}} \neq P_{A^{\mathcal{FE}}}$ neither.) Thus if both Alice and Bob revised their beliefs in the chance of rain by consulting the

[1] In the calculations we assumed that probabilities of the elements of the extended space that do *not* belong to the original space are determined by their respective Bayesian conditionings; for a related discussion on maximal Bayesian learning rules see the end of Section 4 and Section 5.

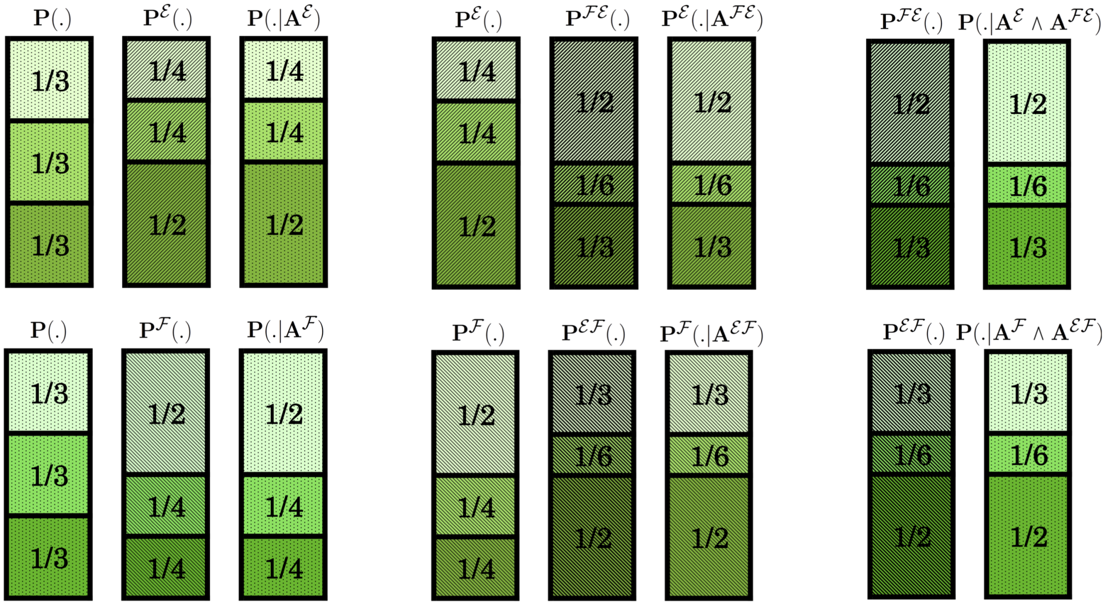

Figure 4: Various probabilities of O_1, O_2, O_3 assuming that the learning rules are maximal Bayesian.

cloudiness of the sky they could not have seen the same clouds; i.e. the sky Alice saw must have had more clouds than the one Bob saw. This is also indicated by the fact that Bob performed a larger revision of belief in the chance of rain (from 3/4 to 1/2) on the basis of his experience than Alice performed (from 2/3 to 1/2) on the basis of hers.

This non-identifiability of the Bayesian factors that induce non-commutative Jeffrey conditionings does not hinge crucially upon the specific extension and upon the specific propositions in the extension we considered. Note first that successively Bayesian conditioning on factors with which we obtained their respective Jeffrey conditionings do commute as expected: for all H in the original probability space

$$P^{\mathcal{FE}}(H) = P_{A^{\mathcal{E}}}(H|A^{\mathcal{FE}}) = P_{A^{\mathcal{FE}}}(H|A^{\mathcal{E}}), \tag{11}$$

$$P^{\mathcal{EF}}(H) = P_{A^{\mathcal{F}}}(H|A^{\mathcal{EF}}) = P_{A^{\mathcal{EF}}}(H|A^{\mathcal{F}}). \tag{12}$$

More importantly by conditioning on the conjunction of these Bayesian factors we also obtain the result of the successive Jeffrey conditioning in our example: for all H in the original probability space

$$P^{\mathcal{FE}}(H) = P_{A^{\mathcal{E}} \wedge A^{\mathcal{FE}}}(H) \tag{13}$$

$$P^{\mathcal{EF}}(H) = P_{A^{\mathcal{F}} \wedge A^{\mathcal{EF}}}(H). \tag{14}$$

If $A^{\mathcal{E}}$ equalled $A^{\mathcal{E}\mathcal{F}}$ and $A^{\mathcal{F}}$ equalled $A^{\mathcal{F}\mathcal{E}}$ then property (13)-(14) would entail $P^{\mathcal{F}\mathcal{E}}(H) = P^{\mathcal{E}\mathcal{F}}(H)$ for all H in the original probability space, contradicting non-commutativity.

Along the same lines, when we have *any* example of finite Jeffrey conditionings that do not commute, and when we have *any* extension with *any* Bayesian factors satisfying (7)-(10) that have the property (13)-(14), these Bayesian factors can not be pairwise identified. Thus Alice and Bob could not have experienced the same pairs of things, no matter what their experience was.

This result also naturally generalizes to the update rules of Field, Gallow etc. Section 5 is going to shed light on the perplexing entré property (13)-(14) made in this discussion.

4 The updating and the updated

One who has reservations about embracing both aspects of probabilistic learning (probability kinematics and proposition kinematics) may consider the following reformulation of the results more illuminating.

As we mentioned in the introduction we can conceptually distinguish between

(i) a set of propositions $\mathcal{L}$, and

(ii) a set of representations of evidences $\mathcal{S}$ on the basis of which the agent may update her subjective beliefs in propositions $\mathcal{L}$.

The basic Bayesian approach equates these two elements: $\mathcal{L} = \mathcal{S}$. This representational choice is, however, rather restrictive. Learning the truth of a proposition may clearly count as evidence for updating subjective beliefs in other propositions; however not all evidence on the basis of which subjective beliefs in propositions can be updated need to come in the form of a proposition. In other words, it seems reasonable to assume that $\mathcal{L}$ forms a part of $\mathcal{S}$, however there are reasons to assume that $\mathcal{S}$ also contains many more elements that are lying outside of $\mathcal{L}$. (Cf. the discussion in Chapter 11.2 of Jeffrey (1983).)

One can interpret $\mathcal{S}$ in different ways. $\mathcal{S}$ could be entailed by a detailed physical-psychological theory of the agent and her possible interactions with her environment. Alternatively, $\mathcal{S}$ could also represent the set of physically possible worlds. Either way it is reasonable to assume that $\mathcal{S}$ is much richer than the set of propositions of a language that the agent is able to formulate.

Definition 4.1 *Let us call a triple* $(\mathcal{L}, \mathcal{S}, P)$ *a* learning frame *if there exists a* $\bar{P}$ *probability measure on* $\mathcal{L}$ *such that* $(\mathcal{S}, P)$ *is an extension of* $(\mathcal{L}, \bar{P})$.

The archetypical example of a learning frame is when $\mathcal{L}$ is a sub-algebra of $\mathcal{S}$, P is a probability measure on $\mathcal{S}$, and $\bar{P} = P_{|\mathcal{L}}$. Thus when $H \in \mathcal{L}$ then also $H \in \mathcal{S}$, and we can simply write $P(H)$ instead of $\bar{P}(H)$. In the spirit of this example in this section we are going to simplify notation: we assume that the homomorphism $\tilde{}$ between $\mathcal{L}$ and $\mathcal{S}$ is fixed; when $H \in \mathcal{L}$ then we identify H with $\tilde{H}$ and simply write $H \in \mathcal{S}$ (instead of writing $\tilde{H} \in \mathcal{S}$); conversely when $H \in \mathcal{S}$ and there exists an $\bar{H} \in \mathcal{L}$ such that $H = \tilde{\bar{H}}$ then we identify H with $\bar{H}$ and simply write $H \in \mathcal{L}$ and $P(H)$ instead of $\bar{P}(\bar{H})$ etc.

Definition 4.2 *A learning frame* $(\mathcal{L}, \mathcal{S}, P)$ *is*

- basic *if* $\mathcal{L} = \mathcal{S}$.
- regular *if*
  - P *is* non-atomic on $\mathcal{S}$*: for any* $A \in \mathcal{S}$ *with* $P(A) \neq 0$ *there exists a* $B \in \mathcal{S}$*,* $B \subseteq A$*,* $P(B) \neq 0$ *such that* $P(B) < P(A)$*.*
  - P *is* atomic on $\mathcal{L}$*: for any* $A \in \mathcal{L}$ *with* $P(A) \neq 0$ *there exists a* $B \in \mathcal{L}$*,* $B \subseteq A$*,* $P(B) \neq 0$ *such that for all* $C \in \mathcal{L}$*,* $C \subsetneq B$*:* $P(C) = 0$*.*

Example of a regular learning frame: let Ω contain countably many sentences of a language, let $\mathcal{L}$ be the smallest Boolean σ-algebra containing elements of Ω, and let $\bar{P}$ be a probability measure on $\mathcal{L}$. There always exists an extension $(\mathcal{S}, P)$ of $(\mathcal{L}, \bar{P})$ such that P is non-atomic on $\mathcal{S}$. Then $(\mathcal{L}, \mathcal{S}, P)$ is a regular learning frame. Also, whenever $(\mathcal{L}, \mathcal{S}, P)$ is a regular learning frame and $P(A) \neq 0$ for an $A \in \mathcal{S}$ then $(\mathcal{L}, \mathcal{S}, P_A)$ is also a regular learning frame.

Definition 4.3 *A* learning rule *is a mapping between learning frames.*
A learning rule $(\mathcal{L}, \mathcal{S}, P_1) \mapsto (\mathcal{L}, \mathcal{S}, P_2)$ *is*

- Bayesian *if there exists an* $A \in \mathcal{S}$ *such that for all* $H \in \mathcal{L}$*:*

$$P_2(H) = P_1(H|A).$$

- (finite) Jeffrey *if there exists a (finite) partition* $\{E_i\}_i$*,* $E_i \in \mathcal{S}$*,* $P_1(E_i) > 0$ *and* $q_i \geq 0$*,* $\sum_i q_i = 1$ *such that for all* $H \in \mathcal{L}$*:*

$$P_2(H) = \sum_i q_i \cdot P_1(H|E_i).$$

- (finite) Gallow 1 *if there exists a (finite) partition* $\{T_i\}_i$, $T_i \in \mathcal{S}$, $P_1(T_i) > 0$, $E_i \in \mathcal{S}$, $P_1(T_iE_i) > 0$ *such that for all* $H \in \mathcal{L}$*:*
$$P_2(H) = \sum_i P_1(H|T_iE_i) \cdot P(T_i).$$
- (finite) Gallow 2 *if there exists a (finite) partition* $\{T_i\}_i$, $T_i \in \mathcal{S}$, $P_1(T_i) > 0$, $E_i \in \mathcal{S}$, $P_1(T_iE_i) > 0$, $\Delta_i > 0$ *such that for all* $H \in \mathcal{L}$*:*
$$P_2(H) = \sum_i P_1(H|T_iE_i) \cdot P(T_i) \cdot \Delta_i.$$
- Field *if there exists a finite partition* $\{E_i\}_{i=1}^{2^n}$, $E_i \in \mathcal{S}$, $P_1(E_i) > 0$, *and* $0 < q_i < 1$, $\sum_{i=1}^{2^n} q_i = 1$ *such that for all* $H \in \mathcal{L}$*:*
$$P_2(H) = \frac{\sum_{i=1}^{2^n} e^{\alpha_i} \cdot P_1(HE_i)}{\sum_{i=1}^{2^n} e^{\alpha_i} \cdot P_1(E_i)}$$
where $\alpha_i = \frac{1}{2^n} \prod_{j=1}^{2^n} \frac{P_2(E_i)}{P_1(E_i)} / \frac{P_2(E_j)}{P_1(E_j)}$ *for all* $i = 1, ..., 2^n$.

A few more technical concepts relating to learning rules:

Definition 4.4 *A learning rule* $(\mathcal{L}, \mathcal{S}, P_1) \mapsto (\mathcal{L}, \mathcal{S}, P_2)$ *is*

- basic *if* $(\mathcal{L}, \mathcal{S}, P_1)$ *is basic.*
- regular *if* $(\mathcal{L}, \mathcal{S}, P_1)$ *is regular.*
- conservative *if* $supp(P_2) \subseteq supp(P_1)$.
- bounded *if there exists a number* $\alpha \geq 1$ *such that for all* $H \in \mathcal{L}$*:*
$$P_2(H) \leq \alpha \cdot P_1(H). \tag{15}$$

Every bounded learning rule is clearly conservative, but the converse is not true.

One can show the following:

Proposition 4.1 *Every bounded regular learning rule is Bayesian.*

Corollary 4.1 *A regular learning rule is*

- *Bayesian if and only if it is finite Jeffrey,*
- *Bayesian if and only if it is finite Gallow 1,*
- *Bayesian if and only if it is finite Gallow 2,*
- *Bayesian if it is Field.*

Under the assumptions Corollary 4.1 shows that we don't need to continue extending our space ad indefiniendum if we want to recover any arbitrary finite Jeffrey conditioning on $\mathcal{L}$ as a Bayesian conditioning, since a single set of representations of evidences can contain all the Bayesian factors needed for recovering any finite Jeffrey conditioning on $\mathcal{L}$.

The same set of representations of evidences is also rich enough to account for non-finite Jeffrey conditionings in an approximate sense, as we are now going to show.

Definition 4.5 *A learning rule* $(\mathcal{L}, \mathcal{S}, P_1) \mapsto (\mathcal{L}, \mathcal{S}, P_2)$ *is* approximate Bayesian *if for all* $i \in \mathbb{N}$ *there exists a subset* $\mathcal{L}_i \subseteq \mathcal{L}$*,* $P_1(\bigvee_{H \in \mathcal{L}_i} H) \to 0$ *as* $i \to \infty$*, and an* $A_i \in \mathcal{S}$ *such that*

- *for all* $H \in \mathcal{L} \backslash \mathcal{L}_i$*:*
$$P_2(H) = P_1(H|A_i),$$
- $\mathcal{L}_{i+1} \subseteq \mathcal{L}_i$*,* $A_{i+1} \subseteq A_i$*.*

Every approximate Bayesian learning rule is Bayesian (as can be seen by setting $\mathcal{L}_i = \emptyset$), but the converse does not hold.

Proposition 4.2 *Every conservative regular learning rule is approximate Bayesian.*

Corollary 4.2 *If a regular learning rule is either*

- *Jeffrey,*
- *Gallow 1,*
- *Gallow 2,*
- *Field,*

then it is approximate Bayesian.

Proposition 4.2 and Corollary 4.2 indicates that the limitation of Bayesian conditioning versus non-finite Jeffrey conditioning stems not from the relative weakness of Bayesian conditioning as a means of updating probabilities, but from the lack of an appropriate account of Bayesian conditioning on sets of measure zero. One can think of Definition 4.5 as providing such an account.[2]

[2] Mathematically a result similar to the extension-based version of Proposition 4.2 can be reached by an alternative approach that allows for conditioning propositions upon elements that belong to a set of elements of an extended space (allowing that different propositions get conditioned on different elements from the set). Such approach was proposed by Z. Gyenis and M. Rédei during the first workshop of the Budapest-Krakow Research Group on Probability,

Note that in the definition of a Bayesian, Jeffrey (Gallow, Field etc.) learning rules we only required the probability updates to work in a certain way for all $H \in \mathcal{L}$, and we stayed silent about how these learning rules should update the probability for other $G \in \mathcal{S} \setminus \mathcal{L}$. When learning rules follow the same probability update formulas for all $G \in \mathcal{S} \setminus \mathcal{L}$ as they do for all $H \in \mathcal{L}$, we may call them *maximal*, i.e. maximal Bayesian, maximal Jeffrey, etc.

If $(\mathcal{L}, \mathcal{S}, P)$ can assumed to be regular Corollary 4.1 suggests the following model of Bayesian learning: the agent always updates her subjective beliefs in propositions $\mathcal{L}$ by conditioning on an $\hat{A} \in \mathcal{S}$ that she learns with certainty. Now, if what indeed triggers the change in the agent's subjective beliefs is learning $\hat{A}$ with certainty, then it seems reasonable to require the "true" learning rule to be maximal Bayesian, that is, to require that the agent updates her probabilities of all elements in $\mathcal{S}$ by conditioning on $\hat{A}$. From this it follows, however, that the same learning rule can in general only be Jeffrey, but not maximal Jeffrey (in the non-trivial sense that excludes the case when the maximal Jeffrey rule is also maximal Bayesian, that is when we Jeffrey condition on partition with an event with posterior probability one).

5 Are not maximal Bayesian learning rules viable?

Are not maximal Bayesian learning rules viable? For instance, is it reasonable to assume that an agent's subjective belief revisions are occasionally best modeled by a (non-trivially) maximal Jeffrey learning rule?

The argument we gave in Section 4 against the viability of maximal Jeffrey learning rules – that is, against learning rules that update all elements of $\mathcal{S}$ via Jeffrey conditioning – is based on accepting Bayesian conditioning as the "true" background evidence assimilating procedure. One could insist, however, that maximal Jeffrey learnings do indeed happen. In the rest of this section we show that regular learning rules that are not maximal Bayesian yield some paradoxical consequences.

We return to our example from Section 3. We consider the same original space, the same extended space, and same elements in the extended space, but for the updated probabilities we assume that they are maximal Jeffrey, that is on all elements of the extended space their values are determined

Causality and Determinism in 2014. However the learning model that could motivate that mathematical result does not seem to me well motivated. (The manuscript is being developed under the provisional title *How much can a Bayesian agent learn?*)

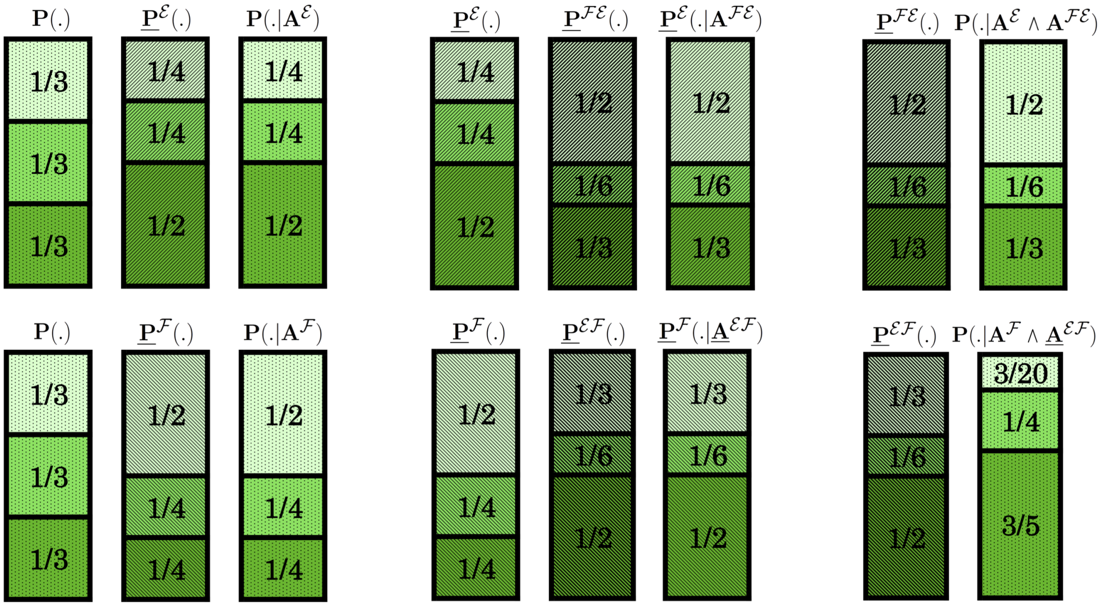
$\mathbf{P}(.)$
1/3
1/3
1/3
$\underline{\mathbf{P}}^{\mathcal{E}}(.)$
1/4
1/4
1/2
$\mathbf{P}(.|\mathbf{A}^{\mathcal{E}})$
1/4
1/4
1/2
$\underline{\mathbf{P}}^{\mathcal{E}}(.)$
1/4
1/4
1/2
$\underline{\mathbf{P}}^{\mathcal{FE}}(.)$
1/2
1/6
1/3
$\underline{\mathbf{P}}^{\mathcal{E}}(.|\mathbf{A}^{\mathcal{FE}})$
1/2
1/6
1/3
$\underline{\mathbf{P}}^{\mathcal{FE}}(.)$
1/2
1/6
1/3
$\mathbf{P}(.|\mathbf{A}^{\mathcal{E}} \wedge \mathbf{A}^{\mathcal{FE}})$
1/2
1/6
1/3
$\mathbf{P}(.)$
1/3
1/3
1/3
$\underline{\mathbf{P}}^{\mathcal{F}}(.)$
1/2
1/4
1/4
$\mathbf{P}(.|\mathbf{A}^{\mathcal{F}})$
1/2
1/4
1/4
$\underline{\mathbf{P}}^{\mathcal{F}}(.)$
1/2
1/4
1/4
$\underline{\mathbf{P}}^{\mathcal{EF}}(.)$
1/3
1/6
1/2
$\underline{\mathbf{P}}^{\mathcal{F}}(.|\underline{\mathbf{A}}^{\mathcal{EF}})$
1/3
1/6
1/2
$\underline{\mathbf{P}}^{\mathcal{EF}}(.)$
1/3
1/6
1/2
$\mathbf{P}(.|\mathbf{A}^{\mathcal{F}} \wedge \underline{\mathbf{A}}^{\mathcal{EF}})$
3/20
1/4
3/5

Figure 5: Various probabilities of O_1, O_2, O_3 assuming that the learning rules are maximal Jeffrey.

by their respective Jeffrey conditioning. We signal this difference by underlining the updated probabilities.

It turns out (see Figure 3 and Figure 5 for details) that we can obtain the $\underline{P}^{\mathcal{E}}$, $\underline{P}^{\mathcal{FE}}$, $\underline{P}^{\mathcal{F}}$, and $\underline{P}^{\mathcal{EF}}$ Jeffrey updated probabilities of Alice and Bob from P by Bayesian conditioning with the same extension. For the same $A^{\mathcal{E}}, A^{\mathcal{FE}}, A^{\mathcal{F}} \in \mathcal{L}$, and for a new element $\underline{A}^{\mathcal{EF}}$ we have, for all H in the original space:

$$\underline{P}^{\mathcal{E}}(H) = P_{A^{\mathcal{E}}}(H) \tag{16}$$

$$\underline{P}^{\mathcal{FE}}(H) = \underline{P}^{\mathcal{E}}_{A^{\mathcal{FE}}}(H) \tag{17}$$

$$\underline{P}^{\mathcal{F}}(H) = P_{A^{\mathcal{F}}}(H) \tag{18}$$

$$\underline{P}^{\mathcal{EF}}(H) = \underline{P}^{\mathcal{F}}_{\underline{A}^{\mathcal{EF}}}(H). \tag{19}$$

(We needed to replace $A^{\mathcal{EF}}$ with another element $\underline{A}^{\mathcal{EF}}$ because $\underline{P}^{\mathcal{EF}}(H)$ does not equal $\underline{P}^{\mathcal{F}}_{A^{\mathcal{EF}}}(H)$.)

All seems well. However the example also shows that something strange is going on. We can obtain $\underline{P}^{\mathcal{E}}$ from P by conditioning on $A^{\mathcal{E}}$, we can obtain $\underline{P}^{\mathcal{FE}}$ from $\underline{P}^{\mathcal{E}}$ by conditioning on $A^{\mathcal{FE}}$, and so it seems natural to assume that we can obtain $\underline{P}^{\mathcal{FE}}$ from P by conditioning on the conjunction $A^{\mathcal{E}} \wedge A^{\mathcal{FE}}$. In other words, similarly to the maximal Bayesian case, one could expect that for all H in the original space:

$$\underline{P}^{\mathcal{FE}}(H) = P_{A^{\mathcal{E}} \wedge A^{\mathcal{FE}}}(H) \tag{20}$$

and similarly

$$\underline{P}^{\mathcal{E}\mathcal{F}}(H) = P_{A^{\mathcal{F}} \wedge \underline{A}^{\mathcal{E}\mathcal{F}}}(H) \tag{21}$$

holds, where i.e. (20) is ostensibly derived as

$$\begin{aligned} \underline{P}^{\mathcal{F}\mathcal{E}}(H) &= \underline{P}^{\mathcal{E}}_{A^{\mathcal{F}\mathcal{E}}}(H) = \underline{P}^{\mathcal{E}}(H|A^{\mathcal{F}\mathcal{E}}) = \frac{\underline{P}^{\mathcal{E}}(H \wedge A^{\mathcal{F}\mathcal{E}})}{\underline{P}^{\mathcal{E}}(A^{\mathcal{F}\mathcal{E}})} && (22) \\ &= \frac{P_{A^{\mathcal{E}}}(H \wedge A^{\mathcal{F}\mathcal{E}})}{P_{A^{\mathcal{E}}}(A^{\mathcal{F}\mathcal{E}})} = \frac{P(H \wedge A^{\mathcal{F}\mathcal{E}}|A^{\mathcal{E}})}{P(A^{\mathcal{F}\mathcal{E}}|A^{\mathcal{E}})} = \frac{\frac{P(H \wedge A^{\mathcal{F}\mathcal{E}} \wedge A^{\mathcal{E}})}{P(A^{\mathcal{E}})}}{\frac{P(A^{\mathcal{F}\mathcal{E}} \wedge A^{\mathcal{E}})}{P(A^{\mathcal{E}})}} && (23) \\ &= \frac{P(H \wedge A^{\mathcal{F}\mathcal{E}} \wedge A^{\mathcal{E}})}{P(A^{\mathcal{F}\mathcal{E}} \wedge A^{\mathcal{E}})} = P(H|A^{\mathcal{E}} \wedge A^{\mathcal{F}\mathcal{E}}) = P_{A^{\mathcal{E}} \wedge A^{\mathcal{F}\mathcal{E}}}(H) && (24) \end{aligned}$$

by using (17) for the first and (16) for the fourth equality. (16) however can not be applied for the fourth equality as it was only guaranteed by our construction for elements of the original space, of which $A^{\mathcal{F}\mathcal{E}}$ is not a member.

As it happens (20) does hold in our example, but its counterpart, (21) does not: i.e. $\underline{P}^{\mathcal{E}\mathcal{F}}(O_1) = 1/3$ while $\underline{P}_{A^{\mathcal{F}} \wedge \underline{A}^{\mathcal{E}\mathcal{F}}}(O_1) = 3/20$![3]

Thus even though successively Bayesian conditioning on factors with which we can obtain their respective Jeffrey conditionings do commute as expected, it is not guaranteed that by conditioning on the conjunction of these Bayesian factors we can obtain the result of successive Jeffrey conditionings when the learning rule is not maximal Bayesian! Commutativity (11) and invariance upon conditioning on conjunctions (20) are thus separate properties: when the set of things that we can update are the same as the set of things that can do the updating they both hold, but the latter does not necessarily follow from the former when these two sets of things do not coincide. Commutativity captures the idea that it shouldn't matter whether we receive $A^{\mathcal{E}}$ first and $A^{\mathcal{F}\mathcal{E}}$ second or we receive $A^{\mathcal{F}\mathcal{E}}$ first and $A^{\mathcal{E}}$ second. But it also shouldn't matter whether we receive $A^{\mathcal{E}}$ and $A^{\mathcal{F}\mathcal{E}}$ successively or at the same time (meaning that we receive their conjunction $A^{\mathcal{E}} \wedge A^{\mathcal{F}\mathcal{E}}$), which is what invariance upon conditioning on conjunctions expresses. Commutativity and invariance upon conditioning on conjunctions are often meshed together since they both hold in the basic Bayesian model, but they express different, albeit equally important desiderata about learning models.

The appearance of failure of invariance upon conditioning on conjunctions can be alleviated to some degree:

[3] There does exist an $\underline{A}^* \in \mathcal{L}$ – depicted in Figure 3 – such that for all H in the original space:

$$\underline{P}^{\mathcal{E}\mathcal{F}}(H) = P_{\underline{A}^*}(H), \tag{25}$$

but $\underline{A}^* \neq A^{\mathcal{F}} \wedge \underline{A}^{\mathcal{E}\mathcal{F}}$.

Proposition 5.1 *Let $(\mathcal{L}, \mathcal{S}, P_0)$ be a regular learning frame, for all $k = 1, ..., N$ let $\{E_1^k, ..., E_{n_k}^k\}$ be a finite partition of $\mathcal{S}$, for all $i = 1, ..., n_k$ let $P_0(E_i^k) > 0$, $q_i^k \geq 0$, $\sum_{i=1}^{n_k} q_i^k = 1$, and let $(\mathcal{L}, \mathcal{S}, P_k)$ be regular learning frames such that for all $H \in \mathcal{S}$:*

$$P_k(H) = \sum_{i=1}^{n_k} q_i^k \cdot P_{k-1}(H|E_i^k).$$

Then for all $k = 1, ..., N$ there exists an $A_k \in \mathcal{S}$ such that for all $H \in \mathcal{L}$:

$$P_k(H) = P_{k-1}(H|A_k), \tag{26}$$

and for which

$$P_k(H) = P_0(H|\bigwedge_{i=1}^{k} A_i). \tag{27}$$

Even thought Proposition 5.1 shows that we can always obtain successive maximal finite Jeffrey conditionings by successive Bayesian conditionings on factors in a way that by conditioning on the conjunction of these factors we can also obtain the result of the successive maximal finite Jeffrey conditionings, there is something deeply disturbing in the construction that is required to achieve this: in order to determine a Bayesian factor at stage k we need to have already determined all other Bayesian factors that will follow *after* stage k. Thus if we want to reconstruct successive maximal finite Jeffrey conditionings as Bayesian conditionings with retaining invariance upon conditioning on conjunctions then we need to require the agent to have foresight in what other Jeffrey conditionings she will perform in the future. This problem may be labeled as the *paradox of future dependence of conditioning on conjunctions for non-maximal Bayesian learning rules.*

The paradox of future dependence of conditioning on conjunctions can only be avoided when the regular learning rule is maximal Bayesian; in this case invariance upon conditioning on conjunctions is also automatically satisfied. (In the maximal Bayesian case there is no future dependence since then condition (33) in the proof of Proposition 5.1 is satisfied by all elements of $\mathcal{S}$, not just those of the form $H \wedge \bigwedge_{i=k+1}^{N} A_i$, and hence there is no dependence on what A_i, $i = k+1, ..., N$ are.)

6 A disjunctive model of Bayesian learning

One may insist that there are cases when the agent only learns new probability values q_i on a partition $\mathcal{E}$ and updates her subjective beliefs without learning anything with certainty. Taken literally the existence of such cases does not strike me plausible. Even if the agent receives the new information i.e. on a slip of paper, there have been a change in the interaction of the agent and the physical world which can be modeled by the agent learning something with certainty, i.e. that she had the experience of reading this-and-that on a slip of paper. If $\mathcal{S}$ is rich enough to represent such physical interactions and experiences then successive Bayesian conditioning on elements of $\mathcal{S}$ remains an adequate model of updating subjective beliefs.

This is not to say that we shouldn't want to model situations in which either the proper source of information – the specific $\hat{A} \in \mathcal{S}$ which represents the physical interaction that triggers the change of beliefs – is uninteresting for the agent, or in which it is impractical or unfeasible or uninteresting to construct the detailed physical-psychological theory that models the information interactions of the agent. There are clearly many pragmatic reasons why we may want to rely on restricted models that do not take these details into account. These pragmatic concerns can however be accommodated without giving up Bayesian conditioning as the core model of subjective belief revision. We can easily incorporate into the Bayesian model the lack of specification of the $\hat{A} \in \mathcal{S}$ that triggers the change of beliefs by tracking not only the single conditional probability distribution that is conditioned on $\hat{A}$ but a set of conditional probability distributions that are conditioned on elements of $\mathcal{S}$ which lead to the same updated probability on $\mathcal{L}$ as does $\hat{A}$.

This suggests the following *disjunctive model of Bayesian learning.* Initially the agent's subjective beliefs about propositions are represented by a probability space $(\mathcal{L}, \bar{P})$ where $\bar{P}$ in non-atomic on $\mathcal{L}$. A detailed physical-psychological theory that models the information interactions of the agent would assigns to $(\mathcal{L}, \bar{P})$ an extension $(\mathcal{S}, P)$ so that $(\mathcal{L}, \mathcal{S}, P)$ is a regular learning frame; we do not know the details of how this extension is obtained, but it is sufficient to assume that it exists. The agent's subjective beliefs at any later stage n are then going to be represented by a triple $(\mathcal{L}, \mathcal{S}, \mathcal{P}_n)$ where $\mathcal{P}_n$ is a set of probability measures defined on $\mathcal{S}$ which all agree with the same $P_n^{\mathcal{L}}$ probability measure defined on $\mathcal{L}$, where $P_0^{\mathcal{L}} = \bar{P}$. Suppose that the agent's beliefs on $\mathcal{L}$ change from $P_n^{\mathcal{L}}$ to $P_{n+1}^{\mathcal{L}}$ such that these probabilities satisfy condition (15). (This change of beliefs may be due to, say, Bayesian, Jeffrey, Gallow, Field etc. sort of conditioning upon a proposition(-partition).) Then

$$\begin{aligned}\mathcal{P}_{n+1} \quad = \quad & \{P' : \exists P_n \in \mathcal{P}_n, \exists A \in \mathcal{S} : \forall G \in \mathcal{S} : P'(G) = P_n(G|A) \\ & \text{and } \forall H \in \mathcal{L} : P'(H) = P_{n+1}^{\mathcal{L}}(H)\}.\end{aligned}$$

This disjunctive model is based purely on Bayesian conditioning yet is able to accommodate a host of other proposed models of subjective belief revision, including Jeffrey conditioning. It assumes that changes in the subjective probability of propositions is always triggered by learning a Bayesian factor with certainty – and so we should think of such Bayesian factors, and not of Jeffrey's partition-value pairs, as representations of incoming evidence –, but it assumes lack of specificity of this Bayesian factor by working with a set of probability distributions that are induced by all possible suitable factors. Thus according to the disjunctive model Jeffrey conditioning is better understood as providing a method for incorporating *unspecified* but *certain* evidence rather than providing a method for incorporating *specific* but *uncertain* evidence.

Departing from the disjunctive model we close this section with a note on the ostensibly problematic *irreversibility* of Bayesian conditioning. Jeffrey conditioning is often touted as superior to Bayesian conditioning since it has the advantage of being reversible: mistakes can be erased (Jeffrey; 1983, p. 172). Indeed an agent should be able to revert a change of belief in a proposition that was triggered by having a specific experience, for her change of belief in the proposition may also depend on background assumptions that influence how said specific experience gets evaluated, and these background assumptions themselves may later change in a way that annuls the effect of said specific experience. Our account respects this requirement: any mistakes that can be erased on $\mathcal{L}$ by Jeffrey or Gallow conditioning can also be erased by an appropriate Bayesian conditioning on an element of $\mathcal{S}$. (Cf. with the criticism Weisberg (2009) mounts against conditionalization on the basis of not being holistic and with the claim of Gallow (2014) that his proposed update rule does abide holism.) However sans memory loss we should not expect the agent to be able to erase the fact that she had the specific experience itself. Thus it is an advantage of our account that both the facts of committing and erasing a mistake gets recorded in changes of probability on $\mathcal{S}$.

Conditioning on a specific $A \in \mathcal{L}$ is indeed irreversible in $\mathcal{L}$. However one wonders how serious this problem is. Typically one wants to think of elements of $\mathcal{L}$ as propositions of a language, i.e. statements of scientific theories. Sans divine intervention no agent is going to learn directly such scientific statements, but only confirm or disconfirm them via observation and experimentation. If we accept the ethos that confirmation and disconfirmation of scientific statements via observation is never absolutely certain, and if we think of $\mathcal{S}$ as containing, among else, the set of representations of observations via which the agents can confirm and disconfirm propositions in $\mathcal{L}$, then any mistake that the agent can commit during her quest to confirm or disconfirm statements of scientific theories can always be erased. And that should be sufficient.

Acknowledgement

I'm grateful for substantial discussions I had with Márton Gömöri, Attila Molnár, Tamás Bitai, Péter Juhász, Gábor Szabó, and for helpful comments from attendees of the second workshop of the Budapest-Krakow Research Group on Probability, Causality, and Determinism, especially those of Leszek Wroński, Zalán Gyenis, and Sam Fletcher.

Appendix

Proof of Proposition 2.1. If P_2 can be obtained from P_1 by Bayesian conditioning without extension then there exists an $A \in \mathcal{L}$, $P_1(A) > 0$ such that for all $H \in \mathcal{L}$ we have $P_2(H) = P_1(H|A)$. If $P_1(A) = 1$ then choose an arbitrary $E \in \mathcal{L}$, $0 < P(E) < 1$ and note that $P_2(H) = P_1(H|A) = P_1(H) = P_1(E) \cdot P_1(H|E) + P_1(\neg E) \cdot (H|\neg E)$ and thus P_2 is obtained by finite Jeffrey conditioning from P_1 without extension using partition $\{E, \neg E\}$ and $q_1 = 1 - q_2 = P_1(E)$. If $P_1(A) \neq 1$ then $P_2(H) = P_1(H|A) = 1 \cdot P_1(H|A) + 0 \cdot P_1(H|\neg A)$, which shows that P_2 is obtained from P_1 by finite Jeffrey conditioning without extension using partition $\{A, \neg A\}$ and setting $q_1 = 1 - q_2 = 1$. □

Proof of Counterexample 2.1. Let $\mathcal{L} = \{\emptyset, a, b, \{a, b\}\}$, $P_1(a) = 1 - P_1(b) = 0.5$, $P_2(a) = 1 - P_2(b) = 0.3$, then P_2 can be obtained from P_1 by finite Jeffrey conditioning without extension by setting $E = \{a\}$ but P_2 can not be obtained from P_1 by Bayesian conditioning without extension, as it can be quickly checked. □

The following is a generalization of Theorem 2.1 of Diaconis and Zabell (1982) that covers probability spaces whose base is not necessarily countable (the proof is essentially the same):

Lemma 1 *P_2 can be obtained from P_1 by Bayesian conditioning with extension if and only if there exists a number $\alpha \geq 1$ such that*

$$P_2(H) \leq \alpha \cdot P_1(H) \tag{28}$$

for all $H \in \mathcal{L}$.

Proof of Lemma 1. If P_2 can be obtained from P_1 by Bayesian conditioning with extension then there exists an extension $(\tilde{\Omega}, \tilde{\mathcal{L}}, \tilde{P}_1)$ of $(\Omega, \mathcal{L}, P_1)$ and an $\hat{A} \in \tilde{\mathcal{L}}$, $\tilde{P}_1(\hat{A}) > 0$ such that for all $H \in \mathcal{L}$: $P_2(H) = \tilde{P}_1(\tilde{H}|\hat{A})$. Then for any $H \in \mathcal{L}$:

$$P_2(H) = \tilde{P}_1(\tilde{H}|\hat{A}) \leq \frac{1}{\tilde{P}_1(\hat{A})} \cdot \tilde{P}_1(H), \tag{29}$$

which shows that (28) holds with $\alpha = \frac{1}{\tilde{P}_1(\hat{A})}$.

On the converse suppose that (28) holds with $\alpha \geq 1$. If $\alpha = 1$ then for all $H \in \mathcal{L}$: $P_2(H) = P_1(H)$ and hence the proposition is obvious with setting $\hat{A} = \Omega$. If $\alpha > 1$ then define

$$P_3(H) = \frac{\alpha}{\alpha - 1} P_1(H) - \frac{1}{\alpha - 1} P_2(H) \tag{30}$$

for all $H \in \mathcal{L}$. P_3 is a probability on $(\Omega, \mathcal{L})$ and $P_1 = \frac{1}{\alpha} P_2 + (1 - \frac{1}{\alpha}) P_3$.

Let $\tilde{\mathcal{L}} = \mathcal{L} \times \{a, b\}$. $\tilde{\mathcal{L}}$ is a σ-algebra that contains elements of the form $\hat{H} = (H_1, a) \vee (H_2, b)$ for some $H_1, H_2 \in \mathcal{L}$. The homomorphism $\tilde{}$ identifies an element $H \in \mathcal{L}$ with $\tilde{H} = (H, a) \vee (H, b) \in \tilde{\mathcal{L}}$. Define $\tilde{P}_1(\hat{H}) = \frac{1}{\alpha} P_2(H_1) + (1 - \frac{1}{\alpha}) P_3(H_2)$ for all $H_1, H_2 \in \mathcal{L}$, $\hat{H} = (H_1, a) \vee (H_2, b)$, and let $\hat{A} = (\Omega, a)$. Then $\tilde{P}_1$ is a probability on $\tilde{\mathcal{L}}$ and for an arbitrary

$H \in \mathcal{L}$ we have $\tilde{P}_1(\tilde{H}|\hat{A}) = \tilde{P}_1((H,a) \vee (H,b)|(\Omega,a)) = \frac{\tilde{P}_1((H,a))}{\tilde{P}_1((\Omega,a))} = \frac{1/\alpha}{1/\alpha} \cdot P_2(H) = P_2(H)$. Hence P_2 can be obtained from P_1 by Bayesian conditioning on $\hat{A}$ with extension $(\tilde{\Omega}, \tilde{\mathcal{L}}, \tilde{P}_1)$. □

Proof of Proposition 2.2. Suppose P_2 can be obtained from P_1 by finite Jeffrey conditioning without extension, and thus let $\{E_i\}_{i=1}^n$ be a finite partition with $P_1(E_i) > 0$, let $P_2(E_i) \geq 0$, $\sum_{i=1}^n P_2(E_i) = 1$, and let P_2 be defined by the Jeffrey formula (1). To show that P_2 can be obtained from P_1 by Bayesian conditioning with extension it is sufficient to show the existence of a number $\alpha \geq 1$ such that

$$\alpha \cdot P_1(H) \geq P_2(H) = \sum_{i=1}^n P_2(E_i) \cdot P_1(H|E_i) \tag{31}$$

for all $H \in \mathcal{L}$, according to Lemma 1. Let $\alpha = \max\{2, \sum_{i=1}^n \frac{P_2(E_i)}{P_1(E_i)}\}$. If $P_1(H) > 0$ then $\alpha \geq \sum_{i=1}^n \frac{P_2(E_i)}{P_1(E_i)} \cdot P_1(E_i|H) = \sum_{i=1}^n \frac{P_2(E_i)}{P_1(H)} \cdot P_1(H|E_i)$ and hence $\alpha \cdot P_1(H) \geq \sum_{i=1}^n P_2(E_i) \cdot P_1(H|E_i) = P_2(H)$. If $P_1(H) = 0$ then $P_2(H) = 0$, and so we can conclude that for all $H \in \mathcal{L}$: $\alpha \cdot P_1(H) \geq P_2(H)$. □

Proof of Counterexample 2.2. Let $\Omega = \{\omega_1, \omega_2, ...\}$ be countable, let $P_1(\omega_i) = \frac{1}{2^i}$, let $\tilde{\Omega} = \Omega \times \{a, b\}$ and $\tilde{\omega}_i = (\omega_i, a) \vee (\omega_i, b)$, let $\tilde{P}_1((\omega_i, a)) = \frac{1}{2^i} P_1(\omega_i) = \frac{1}{4^i}$, let $\hat{A} = \bigvee_i (\omega_i, a)$, and let $P_2(H) = \tilde{P}_1(\tilde{H}|\hat{A})$ for all $H \in \mathcal{L}$. Suppose that P_2 can be obtained from P_1 by (finite or not finite) Jeffrey conditioning without extension with partition $\{E_i\}_i$, $P_1(E_i) > 0$. It follows that for all $\omega_j \in \Omega$ there needs to be an element E_i of this partition such that $\omega_j \in E_i$ and

$$\frac{P_2(\omega_j)}{P_1(\omega_j)} = \frac{P_2(E_i)}{P_1(E_i)} \tag{32}$$

(see Theorem 2.2 of Diaconis and Zabell (1982)). Note that $P_2(\omega_j) = \tilde{P}_1((\omega_j, a) \vee (\omega_j, b)| \bigvee_k (\omega_k, a)) = \frac{\tilde{P}_1((\omega_j,a))}{\tilde{P}_1(\bigvee_k (\omega_k,a))} = \frac{1}{\sum_k \frac{1}{4^k}} \cdot \frac{1}{2^j} P_1(\omega_j) = 3 \cdot \frac{1}{2^j} P_1(\omega_j)$. Thus $\frac{P_2(\omega_j)}{P_1(\omega_j)} = 3 \cdot \frac{1}{2^j}$, which is different for every $j \in \mathbb{N}$, it follows from (32) that the $\{E_i\}_i$ partition contains countably many elements. Hence P_2 can not be obtained from P_1 by *finite* Jeffrey conditioning without extension. □

Proof of Proposition 2.3. Suppose first that P_2 can be obtained from P_1 by Bayesian conditioning with extension, and thus that there exists an extension $(\tilde{\Omega}, \tilde{\mathcal{L}}, \tilde{P}_1)$ of $(\Omega, \mathcal{L}, P_1)$ and an $\hat{A} \in \tilde{\mathcal{L}}$, $\tilde{P}_1(\hat{A}) > 0$ such that for all $H \in \mathcal{L}$: $P_2(H) = \tilde{P}_1(\tilde{H}|\hat{A})$.

If $\tilde{P}_1(\hat{A}) = 1$ then choose an arbitrary $\hat{E} \in \tilde{\mathcal{L}}$, $0 < \tilde{P}_1(\hat{E}) < 1$ and note that $P_2(H) = \tilde{P}_1(\tilde{H}|\hat{A}) = \tilde{P}_1(\tilde{H}) = \tilde{P}_1(\hat{E}) \cdot \tilde{P}_1(\tilde{H}|\hat{E}) + \tilde{P}_1(\neg\hat{E}) \cdot (\tilde{H}|\neg\hat{E})$ and thus P_2 is obtained from P_1 by finite Jeffrey conditioning with extension $(\tilde{\Omega}, \tilde{\mathcal{L}}, \tilde{P}_1)$ using partition $\{\hat{E}, \neg\hat{E}\}$ and $q_1 = 1 - q_2 = \tilde{P}_1(\hat{E})$. If $\tilde{P}_1(\hat{A}) \neq 1$ then $P_2(H) = \tilde{P}_1(\tilde{H}|\hat{A}) = 1 \cdot \tilde{P}_1(\tilde{H}|\hat{A}) + 0 \cdot \tilde{P}_1(\tilde{H}|\neg\hat{A})$, which shows that P_2 is obtained from P_1 by finite Jeffrey conditioning with extension $(\tilde{\Omega}, \tilde{\mathcal{L}}, \tilde{P}_1)$ using partition $\{\hat{A}, \neg\hat{A}\}$ and setting $q_1 = 1 - q_2 = 1$.

Suppose second that P_2 can be obtained from P_1 by finite Jeffrey conditioning with extension, and thus that there exists an extension $(\tilde{\Omega}, \tilde{\mathcal{L}}, \tilde{P}_1)$ of $(\Omega, \mathcal{L}, P_1)$, a partition $\{\hat{E}_i\}_{i=1}^n$ of $\tilde{\Omega}$ with $\tilde{P}_1(\hat{E}_i) > 0$ and $q_i \geq 0$, $\sum_{i=1}^n q_i = 1$ such that for all $H \in \mathcal{L}$: $P_2(H) = \sum_{i=1}^n q_i \cdot \tilde{P}_1(\tilde{H}|\hat{E}_i)$. For an arbitrary $\hat{H} \in \tilde{\mathcal{L}}$ define $\tilde{P}_2(\hat{H}) = \sum_{i=1}^n q_i \cdot \tilde{P}_1(\hat{H}|\hat{E}_i)$, and repeat the proof of Proposition 2.2 applied to $\tilde{P}_1$ and $\tilde{P}_2$. □

Proof of Counterexample 2.3. Let $\Omega = \{\omega_1, \omega_2, ...\}$ be countable, let $P_1(\omega_1) = \frac{5}{6}$, $P_1(\omega_i) = \frac{1}{3^i}$ for $i > 1$, $P_2(\omega_i) = \frac{1}{2^i}$ for $i \geq 1$. Let $E_i = \omega_i$ and let P_2 be defined by the Jeffrey formula (1).

$P(E_k|\omega_j) = 1$ if $k = j$ and $P(E_k|\omega_j) = 0$ otherwise, and thus $\sum_i \frac{P_2(E_i)}{P_1(E_i)} \cdot P_1(E_i|\omega_j) = \frac{P_2(E_j)}{P_1(E_j)} = (3/2)^j$ which goes to infinity as $j \to \infty$. Thus there is no constant $\alpha \geq 1$ such that condition (28) holds for all $\omega_j \in \Omega$, and hence P_2 can not be obtained from P_1 by Bayesian conditioning with extension. □

Proof of Proposition 4.1. Assume that a regular learning rule is bounded, that is there exists an $\alpha \geq 1$ such that for all $H \in \mathcal{L}$ condition (15) holds. We need to show that there exists an $A \in \mathcal{S}$ such that for all $H \in \mathcal{L}$: $P_2(H) = P_1(H|A)$.

When $\alpha = 1$ and hence $P_2(H) = P_1(H)$ the proof is trivial by setting A to be the unit element of $\mathcal{S}$.

Suppose that (15) holds with $\alpha > 1$. Since P_1 is atomic on $\mathcal{L}$ there exists a set of at most countably many pairwise disjoint $O_i \in \mathcal{L}$ which have the property that $P_1(O_i) > 0$ but for all $C \in \mathcal{L}$, $C \subsetneq O_i$: $P_1(C) = 0$. Every $H \in \mathcal{L}$ can be obtained as $H = \bigvee_{i \in I_H} O_i \vee H_0$ where I_H is the set of indexes i such that $O_i \subseteq H$ and where for $H_0 \in \mathcal{L}$: $P_1(H_0) = 0$.

Since due to condition (15) $P_1(O_i) \geq \frac{1}{\alpha} \cdot P_2(O_i)$ for all O_i, and since P_1 is non-atomic on $\mathcal{S}$, for all O_i there exist an $A_i \subseteq O_i$, $A_i \in \mathcal{S}$ such that $P_1(A_i) = \frac{1}{\alpha} \cdot P_2(O_i)$. Let $A = \bigvee_i A_i$. Then for an arbitrary $H \in \mathcal{L}$ we have $P_1(H|A) = \frac{1}{P_1(\bigvee_i A_i)} \cdot P_1((\bigvee_{i \in I_H} O_i \vee H_0) \wedge (\bigvee_i A_i)) = \frac{1}{1/\alpha} \cdot P_1(\bigvee_{i \in I_H} A_i) = \frac{1}{1/\alpha} \cdot \sum_{i \in I_H} P_1(A_i) = \frac{1/\alpha}{1/\alpha} \cdot \sum_{i \in I_H} P_2(O_i) = P_2(\bigvee_{i \in I_H} O_i) = P_2(H)$. □

Proof of Corollary 4.1. The → directions from Bayesian follows immediately from the fact that with the right choice of parameters finite Jeffrey, finite Gallow 1 and finite Gallow 2 reduces to conditioning (c.f. proof of Proposition 2.1).

To ← directions to Bayesian follows from that finite Jeffrey, finite Gallow 1, finite Gallow 2, and Field learning rules are bounded (we showed this for finite Jeffrey in the proof of Proposition 2.2, the rest are analogous.) □

Proof of Proposition 4.2. Let our learning rule $(\mathcal{L}, \mathcal{S}, P_1) \mapsto (\mathcal{L}, \mathcal{S}, P_2)$ be conservative and regular. Since then P_1 is atomic on $\mathcal{L}$ there exists a set of at most countably many pairwise disjoint $O_i \in \mathcal{L}$ which have the property that $P_1(O_i) > 0$ but for all $C \in \mathcal{L}$, $C \subsetneq O_i$: $P_1(C) = 0$. Every $H \in \mathcal{L}$ can be obtained as $H = \bigvee_{i \in I_H} O_i \vee H_0$ where I_H is the set of indexes i such that $O_i \subseteq H$ and where for $H_0 \in \mathcal{L}$: $P_1(H_0) = 0$. Let $\mathcal{O} = \{O_i\}_i$.

For every $\alpha > 1$ let $\mathcal{O}_\alpha = \{O \in \mathcal{O} : P_2(O) \leq \alpha \cdot P_1(O)\}$, $\mathcal{O}_{\epsilon_\alpha} = \mathcal{O} \backslash \mathcal{O}_\alpha$, $\mathcal{L}_{\epsilon_\alpha} = \{H \in \mathcal{L} : H = \bigvee_{O \in \mathcal{O}_{\epsilon_\alpha}} O\}$, and let $\epsilon_\alpha = P_1(\bigvee_{O \in \mathcal{O}_{\epsilon_\alpha}} O) = P_1(\bigvee_{H \in \mathcal{L}_{\epsilon_\alpha}} H)$. Since by conservativeness if $P_2(O) > 0$ then $P_1(O) > 0$, for any $O \in \mathcal{O}$ there exists a large enough α^* so that $O \in \mathcal{O}_\alpha$ for every $\alpha > \alpha^*$, and thus it is clear that $\epsilon_\alpha \to 0$ as $\alpha \to \infty$.

Let us fix now an $\alpha > 1$. There exists a large enough $\beta > \alpha$ such that $\frac{1 - P_2(\bigvee_{O \in \mathcal{O}_\alpha} O)}{\beta} \leq P_1(\bigvee_{O \in \mathcal{O}_{\epsilon_\alpha}} O)$. Since P_1 is non-atomic on $\mathcal{S}$ there exists an $A_{\epsilon_\alpha} \subseteq \bigvee_{O \in \mathcal{O}_{\epsilon_\alpha}} O$, $A_{\epsilon_\alpha} \in \mathcal{S}$ such that $P_1(A_{\epsilon_\alpha}) = \frac{1 - P_2(\bigvee_{O \in \mathcal{O}_\alpha} O)}{\beta}$. Also, since for all $O \in \mathcal{O}_\alpha$: $P_1(O) \geq \frac{1}{\alpha} \cdot P_2(O)$ and hence $P_1(O) \geq \frac{1}{\beta} \cdot P_2(O)$, for all $O \in \mathcal{O}_\alpha$ there exist an $A_O \subseteq O$, $A_O \in \mathcal{S}$ such that $P_1(A_O) = \frac{1}{\beta} \cdot P_2(O)$.

Let $A_\alpha = (\bigvee_{O \in \mathcal{O}_\alpha} A_O) \vee A_{\epsilon_\alpha}$, then $P_1(A_\alpha) = P_1((\bigvee_{O \in \mathcal{O}_\alpha} A_O) \vee A_{\epsilon_\alpha}) = \sum_{O \in \mathcal{O}_\alpha} P_1(A_O) + P_1(A_{\epsilon_\alpha}) = \sum_{O \in \mathcal{O}_\alpha} \frac{1}{\beta} \cdot P_2(O) + \frac{1 - P_2(\bigvee_{O \in \mathcal{O}_\alpha} O)}{\beta} = \frac{1}{\beta} \cdot P_2(\bigvee_{O \in \mathcal{O}_\alpha} O) + \frac{1 - P_2(\bigvee_{O \in \mathcal{O}_\alpha} O)}{\beta} = \frac{1}{\beta}$.

Then for an arbitrary $H \in \mathcal{L} \backslash \mathcal{L}_{\epsilon_\alpha}$ ($H = \bigvee_{i \in I_H} O_i \vee H_0$ with $O_i \in \mathcal{O}_\alpha$ for $i \in I_H$) we have $P_1(H|A_\alpha) = \frac{1}{P_1(A_\alpha)} \cdot P_1((\bigvee_{i \in I_H} O_i \vee H_0) \wedge ((\bigvee_{O \in \mathcal{O}_\alpha} A_O) \vee A_{\epsilon_\alpha})) = \frac{1}{1/\beta} \cdot P_1(\bigvee_{i \in I_H} A_{O_i}) = \frac{1}{1/\beta} \cdot \sum_{i \in I_H} P_1(A_{O_i}) = \frac{1/\beta}{1/\beta} \cdot \sum_{i \in I_H} P_2(O_i) = P_2(\bigvee_{i \in I_H} O_i) = P_2(H)$.

Finally note that in the construction A_α and $\mathcal{L}_{\epsilon_\alpha}$ can be chosen such that $A_{\alpha^*} \subseteq A_\alpha$ and $\mathcal{L}_{\epsilon_{\alpha^*}} \subseteq \mathcal{L}_{\epsilon_\alpha}$ whenever $\alpha^* \geq \alpha$. □

Proof of Corollary 4.2. This follows from Proposition 4.2 and the fact that Jeffrey, Gallow 1, Gallow 2 and Field learning rules are conservative. □

Proof of Proposition 5.1. We only need to show that the proof of Proposition 4.1 can be carried out so that the resulting set of A_ks respect condition (27) for all $k = 1, ..., N$ and for all $H \in \mathcal{L}$. For this it is sufficient (and necessary) to guarantee that for all $H \in \mathcal{L}$ all $\overset{?}{=}$ holds as equality in

$$\begin{aligned}
P_N(H) &= P_{N-1}(H|A_N) = \frac{P_{N-1}(H \wedge A_N)}{P_{N-1}(A_N)} \\
&\overset{?}{=} \frac{P_{N-2}(H \wedge A_N|A_{N-1})}{P_{N-2}(A_N|A_{N-1})} = \frac{P_{N-2}(H \wedge A_N \wedge A_{N-1})}{P_{N-2}(A_N \wedge A_{N-1})} \\
&\overset{?}{=} \frac{P_{N-3}(H \wedge A_N \wedge A_{N-1}|A_{N-2})}{P_{N-2}(A_N \wedge A_{N-1}|A_{N-2})} = \frac{P_{N-3}(H \wedge A_N \wedge A_{N-1} \wedge A_{N-2})}{P_{N-2}(A_N \wedge A_{N-1} \wedge A_{N-2})} \\
&\overset{?}{=} ... \\
&\overset{?}{=} \frac{P_1(H \wedge \bigwedge_{i=3}^N A_i|A_2)}{P_1(\bigwedge_{i=3}^N A_i|A_2)} = \frac{P_1(H \wedge \bigwedge_{i=2}^N A_i)}{P_1(\bigwedge_{i=2}^N A_i)} \\
&\overset{?}{=} \frac{P_0(H \wedge \bigwedge_{i=2}^N A_i|A_1)}{P_0(\bigwedge_{i=2}^N A_i|A_1)} = \frac{P_0(H \wedge \bigwedge_{i=1}^N A_i)}{P_0(\bigwedge_{i=1}^N A_i)} = P_0(H|\bigwedge_{i=1}^N A_i).
\end{aligned}$$

For this latter it is sufficient to guarantee that

$$P_k(H \wedge \bigwedge_{i=k+1}^N A_i) = P_{k-1}(H \wedge \bigwedge_{i=k+1}^N A_i|A_k) \qquad \forall k = 1, ..., N-1, \quad \forall H \in \mathcal{L}. \tag{33}$$

(33) can be guaranteed as follows: first carry out the construction of A_N that satisfies (26) with $k = N$ by following the proof of Proposition 4.1. Let then $\mathcal{L}_{N-1}$ be the smallest σ-algebra containing $\mathcal{L}$ and A_N. P_{N-1} is also atomic on $\mathcal{L}_{N-1}$ and thus we can carry out the construction of $A_{N-1} \in \mathcal{S}$ by following the proof of Proposition 4.1 so that A_{N-1} satisfies

$$P_{N-1}(G) = P_{N-2}(G|A_{N-1})$$

for all $G \in \mathcal{L}_{N-1}$ (instead merely for all $G \in \mathcal{L}$). Since any $G \in \mathcal{L}_{N-1}$ takes one of the three forms $G = H \wedge A_N$, $G = H \wedge \neg A_N$, or $G = H$ for some $H \in \mathcal{L}$, this way we guaranteed

$$P_{N-1}(H \wedge A_N) = P_{N-2}(H \wedge A_N|A_{N-1})$$

for all $H \in \mathcal{L}$.

Let then $\mathcal{L}_{N-2}$ be the smallest σ-algebra containing $\mathcal{L}_{N-1}$ and A_{N-1}, repeat the procedure above to obtain $A_{N-2} \in \mathcal{S}$ that satisfies

$$P_{N-2}(G) = P_{N-3}(G|A_{N-2})$$

for all $G \in \mathcal{L}_{N-2}$, thereby guaranteeing that

$$P_{N-2}(H \wedge A_N \wedge A_{N-1}) = P_{N-3}(H \wedge A_N \wedge A_{N-1}|A_{N-2})$$

for all $H \in \mathcal{L}$ etc. After $N-1$ repetition we obtain the required set of $A_N, A_{N-1}, ..., A_1 \in \mathcal{S}$ which satisfy condition (33). □

Note that if in the proof of Proposition 5.1 we alter condition (33) with an appropriately chosen λ_k to

$$P_k(H \wedge \bigwedge_{i=k+1}^N A_i) = \lambda_k \cdot P_{k-1}(H \wedge \bigwedge_{i=k+1}^N A_i|A_k) \qquad \forall k = 1, ..., N-1, \quad \forall H \in \mathcal{L}$$

then this altered condition also becomes necessary for guaranteeing condition (27), and proceeding along such an altered condition would still lead to dependence on future factors, as explained in the end of Section 5.

References

Diaconis, P. and Zabell, S. L. (1982). Updating subjective probability, *Journal of the American Statistical Association* **77**(380): 822–830.

Döring, F. (1999). Why bayesian psychology is incomplete?, *Philosophy of Science* **66**: S379–S389.

Field, H. (1978). A note on jeffrey conditionalization, *Philosophy of Science* **45**(3): 361–367.

Gallow, J. D. (2014). How to learn from theory-dependent evidence; or commutativity and holism: A solution for conditionalizers, *The British Journal for the Philosophy of Science* **65**: 493–519.

Garber, G. (1980). Field and jeffrey conditionalization, *Philosophy of Science* **47**(142–145).

Jeffrey, R. C. (1983). *The Logic of Decision*, The University of Chicago Press, Chicago.

Osherson, D. (2002). Order dependence and jeffrey conditionalization. Available at `http://philpapers.org/rec/OSHODA`.

Weisberg, J. (2009). Commutativity or holism? a dilemma for conditionalizers, *The British Journal for the Philosophy of Science* **60**: 793–812.