LSE Research Online

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Zalán Gyenis and Miklós Rédei

General properties of Bayesian learning as statistical inference determined by conditional expectations

Article (Published version) (Refereed)

Original citation:
Gyenis, Zalán and Rédei, Miklós (2017) *General properties of Bayesian learning as statistical inference determined by conditional expectations.* Review of Symbolic Logic. ISSN 1755-0203

DOI: 10.1017/S1755020316000502

GENERAL PROPERTIES OF BAYESIAN LEARNING AS STATISTICAL INFERENCE DETERMINED BY CONDITIONAL EXPECTATIONS

ZALÁN GYENIS

Department of Algebra, Budapest University of Technology and Economics

and

MIKLÓS RÉDEI

Department of Philosophy, Logic and Scientific Method
London School of Economics and Political Science

Abstract. We investigate the general properties of general Bayesian learning, where "general Bayesian learning" means inferring a state from another that is regarded as evidence, and where the inference is conditionalizing the evidence using the conditional expectation determined by a reference probability measure representing the background subjective degrees of belief of a Bayesian Agent performing the inference. States are linear functionals that encode probability measures by assigning expectation values to random variables via integrating them with respect to the probability measure. If a state can be learned from another this way, then it is said to be Bayes accessible from the evidence. It is shown that the Bayes accessibility relation is reflexive, antisymmetric, and nontransitive. If every state is Bayes accessible from some other defined on the same set of random variables, then the set of states is called weakly Bayes connected. It is shown that the set of states is not weakly Bayes connected if the probability space is standard. The set of states is called weakly Bayes connectable if, given any state, the probability space can be extended in such a way that the given state becomes Bayes accessible from some other state in the extended space. It is shown that probability spaces are weakly Bayes connectable. Since conditioning using the theory of conditional expectations includes both Bayes' rule and Jeffrey conditionalization as special cases, the results presented generalize substantially some results obtained earlier for Jeffrey conditionalization.

§1. Review of main results. In this paper we investigate the general properties of general Bayesian learning. The investigation is motivated by the observation that the properties of Bayesian learning we wish to determine do not seem to have been analyzed in the literature on Bayesianism on the level of generality we aim at here. (For monographic works on Bayesianism we refer to [3, 23, 47]; for papers discussing basic aspects of Bayesianism see [10, 11, 19–22, 45, 46].)

In particular, in this paper we take the position that the proper general technical device to perform Bayesian conditioning is the theory of conditional expectations. The concept of conditional expectation was introduced into probability theory by Kolmogorov in 1933 together with his axiomatization of probability theory, which has made probability theory part of measure theory [29] (Doob [5] puts Kolmogorov's work into historical context).

Received: August 24, 2016.
2010 *Mathematics Subject Classification*. 03A10, 03A05, 03A99.
Key words and phrases. Bayesian learning, conditionalization, Bayes accessibility.

doi:10.1017/S1755020316000502

Since Kolmogorov's work, conditioning using the theory of conditional expectations has become standard in mathematics [1, 2, 6, 12, 37, 41]. Both the elementary Bayes' rule (sometimes called "strict conditionalization") and Jeffrey conditionalization (also called "probability kinematics" [26]) can be recovered as special cases of conditioning using the theory of conditional expectations; although, somewhat surprisingly, the fact that Jeffrey conditionalization is indeed a special case of conditioning via conditional expectations does not seem to be well known: Jeffrey does not refer to the theory of conditional expectations when introducing his rule of conditionalization in [26], nor do standard mathematical works on probability theory [1, 2, 12, 37, 41] mention Jeffrey conditionalization when discussing the concept of conditional expectation. With very few exceptions (e.g., [8, 9, 24]) philosophical work on Bayesianism has not made extensive and systematic use of the general theory of conditional expectations as conditioning device.

Proper handling of conditioning via conditional expectation requires one to go beyond the framework of additive measures on Boolean algebras and forces one to work with positive, normalized, linear functionals (called "states") that assign finite expectation values to random variables via integrating them with respect to probability measures on the Boolean algebra. Viewed from this more general perspective, conditioning can be regarded as a map in the state space that takes a state as input and yields another, the conditioned state, as output. Conditioning is thus a map in the dual space of the function space consisting of integrable random variables. *Bayesian* conditioning is distinguished among the logically possible conditioning maps in the dual space by the fact that it is the dual of a specific map (a projection) on the function space representing the random variables that are integrable with respect to the background probability of the Bayesian Agent. This projection on the function space is the conditional expectation. Bayesian conditionalization based on the technique of conditional expectations as conditioning device thus defines a two-place relation in the state space of integrable random variables. We call this relation the "Bayes accessibility relation" (Definition 4.2). The interpretation of the Bayes accessibility relation is that if a state is Bayes accessible from another, then the Bayesian Agent can infer this state from the evidence represented by the other state, where the inference is a single Bayesian conditionalizing using the technique of conditional expectation determined by the Agent's background probability. Characterizing the Bayes accessibility relation amounts then to characterizing Bayesian learning based on a fixed background measure in this general setting.

Note that regarding Bayesian learning this way as a specific type of probabilistic inference based on conditionalization as an inference device amounts to an interpretation of conditionalization that differs slightly from what is common in situations in which conditionalization is not via the most general technique of conditional expectations but via the Bayes or Jeffrey rules. The difference is reflected in our terminology, which differs from the standard one: We say that a state is inferred from another (evidence) on the basis of a background measure via conditionalization determined by the background measure; equivalently: that the evidence is conditionalized using the conditional expectation determined by the background measure (prior) to obtain the inferred, conditioned state (posterior). The standard way of expressing this would be to say that one conditionalizes the background measure (prior) on the evidence to obtain the posterior. We believe that our terminology indicates well the logical aspects and structure of conditionalizing via general conditional expectations. Our terminology also is in harmony with the one used in some approaches to conditionalization in physics (see e.g., [31–33]). More will be said on this in §3, see especially Remark 3.4.

Using the Bayes accessibility relation, one also can define another relation[1]: one that connects a state defined by a probability measure not to the evidence from which it can be learned via conditionalization determined by a fixed background measure but to the background probability measure on the basis of which the (posterior) state can be learned from some evidence (Definition 4.3). We call this relation the "prior-posterior relation". The prior-posterior relation reflects another aspect of general Bayesian learning. Describing properties of the Bayes accessibility and prior-posterior relations amounts to characterizing Bayesian learning based on the concept of conditional expectation as a general inference device. This is what we do in this paper. In the rest of this introduction we review (without using any technical notation) first the results on the Bayes accessibility relation that connects the evidence to the inferred (posterior) probability. The results on the prior-posterior relation are summarized at the end of the introduction.

The Bayes accessibility relation is trivially reflexive—every state can be inferred from itself trivially. The first nontrivial result of the paper is that the Bayes accessibility relation is antisymmetric (Proposition 5.1). Antisymmetry of the Bayes accessibility relation entails that a state space is *not* strongly Bayes connected: It is *not* true that any two states are Bayes accessible from each other. In other words, Bayesian learning has a certain directedness built into it: if a state can be learned from another that represents evidence, then the converse is not true, Bayesian learning in the reversed direction is not possible (assuming that the two states are different). This seems an intuitively attractive feature of Bayesian learning: what one can learn from evidence cannot serve as evidence to learn the evidence itself.

We then ask the question of whether state spaces are *weakly* Bayes connected: is it true that *every* state is Bayes accessible from *some other* state (Definition 6.1)? Weak Bayes connectedness means that for every probability measure there exists some evidence from which the Bayesian Agent can learn that probability by conditionalization with respect to his fixed background degree of belief. Failure of weak Bayes connectedness means that, given the background measure of the Bayesian Agent, there exist states (probability measures) that the Agent cannot learn via Bayesian statistical inference no matter what evidence formulated in the given state space he is presented with. Thus weak Bayes connectedness of the state space would be a sign of strength of Bayesian learning–failure of weak connectedness sets a strong limit to Bayesian learning in the given context. We give a characterization of weak Bayes connectedness of state spaces (Proposition 6.2). This result is used then to show that state spaces of *standard* probability spaces are *not* weakly connected (Propositions 6.7 and 6.8). Standard probability spaces include essentially all the probability spaces that occur in applications of probability theory; in particular, probability spaces with a finite number of random events, and probability theories in which probability is given by a density function with respect to the Lebesgue measure, are standard. In fact, we prove more: in case of a standard probability space there exist an *uncountably infinite* number of probability measures that are inaccessible for the Bayesian Agent (Propositions 6.10 and 6.11). Note that since conditioning now is with respect to conditional expectations, not via the simple Bayes' rule, the existence of Bayes inaccessible states stated by Propositions 6.7, 6.10, and 6.11 has nothing to do with the well known fact that a measure which is obtained from another via the simple Bayes' rule will have to take zero value whenever the prior measure takes on value zero (absolute continuity of conditional probability with respect to prior) and that therefore a lot of measures (for instance all faithful probability measures) cannot be obtained as the result of conditionalizing another

[1] We thank two anonymous referees for suggesting to analyze this relation as well.

measure via the simple Bayes' rule. (A probability measure is faithful if all nonzero events have nonzero probability. (In the philosophy of probability literature faithful probability measures are also called "regular".) A faithful *conditional* probability measure appears in Example 7.2.) The uncountably infinite number of Bayes inaccessible states displaying the failure of weak Bayes connectedness of standard probability spaces are *all* absolutely continuous with respect to the Agent's background probability.

Proposition 6.2, which characterizes weak Bayes connectedness, makes it possible to formulate a condition sufficient to entail that a state space is weakly Bayes connected (Proposition 6.5). Based on this latter condition we give an example of a probability space whose state space is weakly Bayes connected. But the significance for Bayesian learning of the probability space with this weakly Bayes connected state space might be very limited because the cardinality of the Boolean algebra of this weakly Bayes connected state space is much larger than that of the continuum. Thus, only Bayesian Agents capable of comprehending an enormous amount of propositions would be in the position to have degrees of belief in every proposition in such a large set. Whether the notion of Bayesian Agent should include Agents with such extraordinary mental skills, is questionable. In the typical situations when one deals with probabilistic modeling, the concept of a Bayesian Agent with more modest mental powers is sufficient. But in such contexts inaccessibility of certain states via Bayesian inference is the general rule.

Failure of weak Bayes connectedness leads to the question of whether state spaces are weakly Bayes *connectable*: Whether for every state (in particular for a state that is not Bayes accessible from any other state in the given probabilistic framework) there exists a richer probability theory into which the original can be embedded in such a way that the Bayes inaccessible state becomes Bayes accessible from *some* state in the richer framework (Definition 8.2). We show that state spaces are weakly Bayes connectable (Proposition 8.3). This result generalizes the ones obtained by Diaconis and Zabell for the simple Bayes' rule and for Jeffrey conditionalization [4] (also see [14]). Weak Bayes connectability of state spaces means that everything that can be formulated by the Bayesian Agent in terms of a given probability space, can *in principle* be learned by the Agent by Bayesian upgrading—provided that the Agent is allowed to enlarge the propositional basis of his degrees of belief in a consistent manner. We will call this latter upgrading situation "Unlimited Evidence Upgrading" scenario, in contradistinction to the "Limited Evidence Upgrading" situation, in which the evidence available to the Agent is restricted to the set of all states on a given set of random variables. Thus under the Unlimited Evidence Upgrading conditions a Bayesian Agent has a potentially unlimited Bayesian learning capacity.

Weak Bayes connectability of state spaces raises the question of whether state spaces are *strongly* Bayes connectable: whether it is true that any state can be made Bayes accessible from any other by embedding both into a larger state space (Problem 8.6). This problem remains open.

We will also show that the Bayes accessibility relation is *not* transitive (Proposition 7.1). Lack of transitivity of the Bayes accessibility relation makes the following definition of *finite* Bayes accessibility nonredundant: a state is *finitely* Bayes accessible from another if it can be obtained as a result of a finite number of successive conditionalizations from this latter state (Definition 7.3). The interpretation of finite Bayes accessibility is that the Agent can genuinely Bayes-learn from error and feedback: the Agent can learn a (finitely Bayes accessible) state from some evidence in several steps, in each step correcting the state inferred in the preceding step by taking into account information that confirms some of the inferred probabilities as correct, while keeping his background probability fixed. Viewed from the perspective of this learning process, lack of transitivity of Bayes accessibility means that while the Bayesian Agent might be able to learn a state by starting the learning process from an initial evidence and performing several successive steps of conditionalizing based on feedback, the Agent will not in general be able to cut short the learning process by replacing the chain of learning steps by a single Bayesian learning move—"there is no Bayesian royal road to learning".

The notion of finite Bayes accessibility leads naturally to the question of whether state spaces are *finitely* weakly Bayes connected: is it true that *every* state is *finitely* Bayes accessible from *some* other state (Definition 7.4)? Since finite weak Bayes connectedness is a weakening of the simple weak Bayes connectedness, it is easier, in principle, for a state space to be finitely weakly Bayes connected than just being Bayes connected. But a simple argument shows that state spaces that are not weakly Bayes connected are not *finitely* weakly Bayes connected either (Proposition 7.5). This entails that standard probability spaces are not finitely weakly connected (Proposition 7.6). Thus, although a Bayesian Agent can learn much more starting from a given evidence if given feedback and allowed to correct the inferred states in a finite series of successive conditionalizations that uses the feedback, there exist states that remain inaccessible for the Agent no matter from what evidence he starts learning and no matter how many (but finite) times he is given feedback.

It will be seen that the prior-posterior relation is identical to the absolute continuity relation between probability measures. Hence it is reflexive, transitive, and it is not symmetric and not antisymmetric. We will also ask whether the set of probability measures on a σ-algebra is weakly connected with respect to the prior-posterior relation: Is it true that for any probability measure on a σ-algebra there exists another that can serve as a background on the basis of which the Bayesian Agent can learn the probability in a nontrivial way from some evidence? This problem remains open in its full generality. But we isolate conditions under which such a suitable background probability exists (Proposition 4.4). Those sufficient conditions can hold in measure spaces with a finite Boolean algebra, and also in spaces where probabilities are given by a density with respect to the Lebesgue measure. Thus, given a probability measure to be learned in a typical application, there exists a Bayesian Agent with a background probability who can learn the probability measure in question from some evidence on the basis of this background. At the same time, failure of weak and finite weak Bayes connectedness of state spaces shows that this very background makes it impossible for the Agent to learn a lot of other probabilities. This underscores the crucial importance of prior probability in Bayesian learning: A Bayesian Agent always comes equipped with a probability measure that represents the Agent's background beliefs. No matter what this belief is, there will always be an uncountably infinite number of probability measures (all absolutely continuous with respect to the prior of the Agent) that the Agent cannot learn by conditionalization, no matter what evidence he is presented with in the given framework. The set of Bayes inaccessible states is a "blind-spot" of the Bayesian Agent and we regard this "Bayes-blindness" of a Bayesian Agent a serious challenge for Bayesianism.

Existence of nonempty Bayes Blind Spots leads to several questions: One is how presence of nonempty Bayes Blind Spots is compatible with the phenomenon known as "washing out of priors"? This question will be answered in §7 by recalling Earman's [8] formulation of "washing out of priors" in terms of Doob's martingale theorem, and clarifying the relation of the martingale theorem to the concept of finite Bayes accessibility. Other natural questions concern the relationship between Bayes Blind Spots and priors.[2]

[2] We thank an anonymous referee for pointing out such questions.

For instance: Are there states that are in the Bayes Blind Spot of many choices of priors? How are the Bayes Blind Spots of different priors related? We do not know the answers to these questions; these are interesting and nontrivial open problems to investigate further. Another natural question concerns the size of the Bayes Blind Spot determined by a given background probability. The size of the Bayes Blind Spot could be determined in the case of probability spaces with a finite Boolean algebra [16] but the question is open even in case of standard probability spaces.

The structure of the paper is the following. §2 fixes notation and recalls some basic definitions and facts from the theory of conditional expectations. §3 defines conditional probability in terms of conditional expectations and shows how elementary Bayesian upgrading and Jeffrey conditionalization obtain as special cases of conditionalization via conditional expectation. §4 defines the Bayes accessibility relation. §5 proves that the Bayes accessibility relation is antisymmetric and discusses failure of strong Bayes connectedness of state spaces. §6 analyzes weak Bayes connectedness and proves that state spaces of standard probability spaces are not weakly Bayes connected. §7 proves failure of transitivity of Bayes accessibility, defines the finite Bayes accessibility relation and discusses finite weak Bayes connectedness. §8 proves weak Bayes connectability of state spaces. §9 summarizes the main points with some further comments.

§2. Conditional expectations.

We fix some notation that will be used throughout the paper. $(X, \mathcal{S}, p)$ denotes a probability measure space: X is the set of elementary events, $\mathcal{S}$ is a σ-algebra of some subsets of X, p is a probability measure on $\mathcal{S}$. The negation of event $A \in \mathcal{S}$ is denoted by $A^{\perp}$. Given $(X, \mathcal{S}, p)$, $\mathcal{L}^s(X, \mathcal{S}, p)$ denotes the set of $f\colon X \to \mathbb{R}$ measurable functions such that $|f|^s$ is p-integrable. Of special importance are the integrable ($s = 1$), the square-integrable ($s = 2$), and the (essentially) bounded functions, the latter corresponds, formally, to $s = \infty$. Since p is a bounded measure, we have (cf. [42, p. 71], [43])

$$\mathcal{L}^{\infty}(X, \mathcal{S}, p) \subset \mathcal{L}^2(X, \mathcal{S}, p) \subset \mathcal{L}^1(X, \mathcal{S}, p). \tag{1}$$

Identifying functions that are equal except on p-measure zero sets, one obtains the corresponding spaces $L^s(X, \mathcal{S}, p)$ consisting of equivalence classes of functions (notice the notational difference between $\mathcal{L}$ and L). In what follows, in harmony with the usual mathematical practice, we use the same letters f, g etc. to refer to both functions (elements of $\mathcal{L}^s(X, \mathcal{S}, p)$) and equivalence classes of functions (elements of $L^s(X, \mathcal{S}, p)$). The characteristic (indicator) functions χ_A of the sets $A \in \mathcal{S}$ are in $\mathcal{L}^s(X, \mathcal{S}, p)$ for all $A \in \mathcal{S}$.

The probability measure p extends from $\mathcal{S}$ to a linear functional ϕ_p on $\mathcal{L}^s(X, \mathcal{S}, p)$ by the integral:

$$\phi_p(f) \doteq \int_X f\,dp \qquad f \in \mathcal{L}^s(X, \mathcal{S}, p) \tag{2}$$

The value of ϕ_p on a characteristic function χ_A of $A \in \mathcal{S}$ is just the p-probability of A:

$$\phi_p(\chi_A) = \int_X \chi_A\,dp = \int_A dp = p(A) \tag{3}$$

The map $f \mapsto \|f\|_s^2 \doteq \phi_p(|f|^s)$ defines a seminorm $\|\cdot\|_s$ on $\mathcal{L}^s(X, \mathcal{S}, p)$ (only a *semi*norm because in the function space $\mathcal{L}^s(X, \mathcal{S}, p)$ functions differing on p-probability zero sets are *not* identified). The linear functional ϕ_p is continuous in the seminorm $\|\cdot\|_s$. The seminorm $\|\cdot\|_s$ becomes a norm on $L^s(X, \mathcal{S}, p)$. The containment relation (1) is dense when $\mathcal{L}^s$ ($s = 1, 2, \infty$) are considered as normed spaces.

By definition, a state ϕ on $\mathcal{L}^s(X, \mathcal{S}, p)$ is a linear functional on $\mathcal{L}^s(X, \mathcal{S}, p)$ that is positive (i.e., $\phi(f) \geq 0$ if $f \geq 0$), continuous in the $\|\cdot\|_s$ seminorm and normalized $\phi(\mathbf{1}) = 1$, where $\mathbf{1}$ denotes the characteristic function χ_X of the whole set X. States yield probability measures when restricted to the (characteristic functions of elements of the) σ-algebra $\mathcal{S}$. States thus encode probability measures via the integral (2).

REMARK 2.1. *Since states on $\mathcal{L}^s(X, \mathcal{S}, p)$ are $\|\cdot\|_s$-continuous linear functionals by definition, all the probability measures obtained from states on $\mathcal{L}^s(X, \mathcal{S}, p)$ by restricting them to the (characteristic functions of the) σ-algebra $\mathcal{S}$ are absolutely continuous with respect to p. Thus, if $f = g$ almost p-everywhere, then for any state ϕ we have $\phi(f) = \phi(g)$. From this it follows that states on $\mathcal{L}^s(X, \mathcal{S}, p)$ can be regarded as linear functionals on the space $L^s(X, \mathcal{S}, p)$ of equivalence classes of integrable functions. $L^s(X, \mathcal{S}, p)^\sharp$ denotes the set of all states on $L^s(X, \mathcal{S}, p)$. States are a proper subset of the dual space that contain* all *$\|\cdot\|_s$-continuous linear functionals.*

The space of square-integrable random variables $L^2(X, \mathcal{S}, p)$ is a Hilbert space with respect to the scalar product $\langle \cdot, \cdot \rangle$ defined by

$$\langle f, g \rangle \doteq \int_X fg \, dp \qquad f, g \in L^2(X, \mathcal{S}, p) \tag{4}$$

For more details on the above notions (and other mathematical concepts related to L^s-spaces used here without definition) see the standard references for the measure theoretic probability theory [1, 2, 30, 41]. In particular, §19 in [1] and Chapter 3 in [42] discuss further properties of the function spaces $L^s(X, \mathcal{S}, p)$.

The central concept that the modern mathematical theory of conditionalization is based on is the notion of conditional expectation:

DEFINITION 2.2 ([1, p. 445]). *Let $(X, \mathcal{S}, p)$ be a probability space, $\mathcal{A}$ be a σ-subalgebra of $\mathcal{S}$, and $p_{\mathcal{A}}$ be the restriction of p to $\mathcal{A}$. A map*

$$\mathscr{E}(\cdot \mid \mathcal{A}) \colon \mathcal{L}^1(X, \mathcal{S}, p) \to \mathcal{L}^1(X, \mathcal{A}, p_{\mathcal{A}}) \tag{5}$$

is called an $\mathcal{A}$-conditional expectation from $\mathcal{L}^1(X, \mathcal{S}, p)$ to $\mathcal{L}^1(X, \mathcal{A}, p_{\mathcal{A}})$ if (*i*) *and* (*ii*) *below hold*:

(i) *For all $f \in \mathcal{L}^1(X, \mathcal{S}, p)$, the $\mathscr{E}(f \mid \mathcal{A})$ is $\mathcal{A}$-measurable.*

(ii) *$\mathscr{E}(\cdot \mid \mathcal{A})$ preserves the integration on elements of $\mathcal{A}$*:

$$\int_Z \mathscr{E}(f \mid \mathcal{A}) dp_{\mathcal{A}} = \int_Z f \, dp \qquad \forall Z \in \mathcal{A}. \tag{6}$$

The $\mathcal{A}$-measurability condition (i) should be thought of as a coarse-graining requirement: it entails that $\mathscr{E}(f \mid \mathcal{A})$ is constant on minimal elements (atoms) in $\mathcal{A}$ (atoms in $\mathcal{A}$ need not be atoms in $\mathcal{S}$). Condition (ii) is the general form of the theorem of total probability: it requires that from the conditional expectation one can recover the original expectation values (hence the original probability p). For further discussion of the interpretation of properties of the conditional expectation see [1].

It is not obvious that, given a σ-subalgebra $\mathcal{A}$, a conditional expectation $\mathscr{E}(\cdot \mid \mathcal{A})$ exists but the Radon–Nikodym theorem entails that it *always* does:

PROPOSITION 2.3 ([1, p. 445], [2, Theorem 10.1.5]). *Given any $(X, \mathcal{S}, p)$ and any σ-subalgebra $\mathcal{A}$ of $\mathcal{S}$, a conditional expectation $\mathscr{E}(\cdot \mid \mathcal{A})$ from $\mathcal{L}^1(X, \mathcal{S}, p)$ to $\mathcal{L}^1(X, \mathcal{A}, p_{\mathcal{A}})$ exists.*

Note that uniqueness is not part of the claim in Proposition 2.3 because the conditional expectation is only unique up to measure zero:

PROPOSITION 2.4 ([1, Theorem 16.10 and p. 445]; [2, p. 339]). *If $\mathscr{E}'(\cdot \mid \mathcal{A})$ is another conditional expectation then for any $f \in \mathcal{L}^1(X, \mathcal{S}, p)$ the two $\mathcal{L}^1$-functions $\mathscr{E}(f \mid \mathcal{A})$ and $\mathscr{E}'(f \mid \mathcal{A})$ are equal up to a p-probability zero set.*

Different conditional expectations equal up to measure zero are called *versions* of the conditional expectation. It follows that, considered as a map on $L^1(X, \mathcal{S}, p)$, the conditional expectation is unique. We use the notation $\mathbb{E}(\cdot \mid \mathcal{A})$ to denote the conditional expectation $\mathscr{E}(\cdot \mid \mathcal{A})$ when viewed as a map on $L^1(X, \mathcal{S}, p)$.

DEFINITION 2.5 ([1, p. 430]). *Given a conditional expectation $\mathscr{E}(\cdot \mid \mathcal{A})$, the map $\mathscr{P}(\cdot \mid \mathcal{A})$ defined by*

$$\mathcal{S} \ni B \mapsto \mathscr{P}(\chi_B \mid \mathcal{A}) \doteq \mathscr{E}(\chi_B \mid \mathcal{A}) \tag{7}$$

is called the $\mathcal{A}$-conditional probability. $\mathbb{P}(\cdot \mid \mathcal{A})$ denotes the analogue map defined in terms of $\mathbb{E}(\cdot \mid \mathcal{A})$.

Note that the conditional expectation is not the conditional expected value of any random variable—it is a map between function spaces. Nor is the conditional probability a real valued probability measure. Conditional expected values and conditional probabilities as real numbers can be obtained from $\mathscr{E}(\cdot \mid \mathcal{A})$ and $\mathscr{P}(\cdot \mid \mathcal{A})$ (see §3).

The next proposition states some basic features of the conditional expectations.

PROPOSITION 2.6 ([1, sec. 34]). *A conditional expectation has the following properties*:

(i) $\mathbb{E}(\cdot \mid \mathcal{A})$ *is a linear map.*

(ii) $\mathbb{E}(\cdot \mid \mathcal{A})$ *is a projection*:

$$\mathbb{E}(\mathbb{E}(f \mid \mathcal{A}) \mid \mathcal{A}) = \mathbb{E}(f \mid \mathcal{A}) \qquad \forall f \in L^1(X, \mathcal{S}, p) \tag{8}$$

(iii) $\mathbb{E}(\cdot \mid \mathcal{A})$ *preserves the unit*

$$\mathbb{E}(\mathbf{1} \mid \mathcal{A}) = \mathbf{1} \tag{9}$$

(iv) $\mathbb{E}(\cdot \mid \mathcal{A})$ *is a $\|\cdot\|_1$-contraction*: $\|\mathbb{E}(f \mid \mathcal{A})\|_1 \le \|f\|_1$ *(i.e., $\mathbb{E}(\cdot \mid \mathcal{A})$ is continuous in the $\|\cdot\|_1$-norm topology).*

The statements (*i*)–(*iv*) *also hold for $\mathscr{E}(\cdot \mid \mathcal{A})$ except for p-probability zero.*

Note that restricted to the Hilbert space $L^2(X, \mathcal{S}, p)$ the conditional expectation $\mathbb{E}(\cdot \mid \mathcal{A})$ is an orthogonal projection on $L^2(X, \mathcal{S}, p)$ with range $L^2(X, \mathcal{A}, p_{\mathcal{A}})$, a closed linear subspace of $L^2(X, \mathcal{S}, p)$.

A deep result of the theory of conditional expectations is that Properties (i)–(iv) in Proposition 2.6 characterize the conditional expectation completely:

PROPOSITION 2.7 ([37, Theorem 3], [7, Corollary 1], [36]). *Suppose that the map T*

$$T\colon L^1(X, \mathcal{S}, p) \to L^1(X, \mathcal{S}, p) \tag{10}$$

is a linear, $\|\cdot\|_1$-contractive projection preserving $\mathbf{1}$. *Then there exists a σ-subalgebra $\mathcal{A}$ of $\mathcal{S}$ such that T is the conditional expectation from $L^1(X, \mathcal{S}, p)$ to $L^1(X, \mathcal{A}, p_{\mathcal{A}})$.*

§3. Conditional probability in terms of conditional expectation. Let $(X, \mathcal{S}, p)$ be a probability space, $\mathcal{A}$ be a σ-subalgebra of $\mathcal{S}$. Assume that $\psi_{\mathcal{A}}$ is a $\|\cdot\|_1$-continuous linear functional on the subspace $\mathcal{L}^1(X, \mathcal{A}, p_{\mathcal{A}})$ determined by a probability measure $q_{\mathcal{A}}$ given on the subalgebra $\mathcal{A}$ via integral (cf. equation (2)). What is the extension ψ of

$\psi_{\mathcal{A}}$ from $\mathcal{L}^1(X, \mathcal{A}, p_{\mathcal{A}})$ to a $\|\cdot\|_1$-continuous linear functional on $\mathcal{L}^1(X, \mathcal{S}, p)$? This question is the general problem of statistical inference (see [31–33], where the relation of statistical inference to coarse graining also is discussed both in classical and quantum probability theory), and the answer to it is the concept of conditional probability: One is interested in the expectation values $\psi(f)$ of random variables f in $\mathcal{L}^1(X, \mathcal{S}, p)$ that are *not* in $\mathcal{L}^1(X, \mathcal{A}, p_{\mathcal{A}})$ on condition that the expectation values of functions g that *are* in the narrower set of random variables $\mathcal{L}^1(X, \mathcal{A}, p_{\mathcal{A}})$ are prescribed (are known) and are given by $\psi_{\mathcal{A}}(g)$.

In general there are many such extensions. *Bayesian* statistical inference yields a particular answer which is based on Bayesian conditioning via the conditional expectation determined by the probability p and the subalgebra $\mathcal{A}$:

DEFINITION 3.1 (Bayesian statistical inference). *Let the extension ψ of $\psi_{\mathcal{A}}$ be*

$$\psi(f) \doteq \psi_{\mathcal{A}}(\mathscr{E}(f \mid \mathcal{A})) \qquad \forall f \in \mathcal{L}^1(X, \mathcal{S}, p) \tag{11}$$

where $\mathscr{E}(\cdot \mid \mathcal{A})$ is the $\mathcal{A}$-conditional expectation from $\mathcal{L}^1(X, \mathcal{S}, p)$ to $\mathcal{L}^1(X, \mathcal{A}, p_{\mathcal{A}})$.

Recall that $\mathscr{E}(\cdot \mid \mathcal{A})$ is a projection operator on $\mathcal{L}^1(X, \mathcal{S}, p)$ up to p-probability zero (Proposition 2.6) and that state $\psi_{\mathcal{A}}$ is $\|\cdot\|_1$-continuous. Thus ψ is indeed an *extension* of $\psi_{\mathcal{A}}$. Since $\mathscr{E}(\cdot \mid \mathcal{A})$ is $\|\cdot\|_1$-continuous, the extension ψ also is $\|\cdot\|_1$-continuous. Thus equation (11) defines a $\|\cdot\|_1$-continuous extension ψ of $\psi_{\mathcal{A}}$ indeed.

The notion of $(\mathcal{A}, q_{\mathcal{A}})$-conditional *probability* of an event obtains as a special case of Bayesian statistical inference so defined:

DEFINITION 3.2. *If $B \in \mathcal{S}$ then its $(\mathcal{A}, q_{\mathcal{A}})$-conditional probability $q(B)$ is the expectation value $\psi(\chi_B)$ of its characteristic function χ_B computed using the formula (11) containing the $\mathcal{A}$-conditional expectation*:

$$q(B) \doteq \psi(\chi_B) = \psi_{\mathcal{A}}(\mathscr{E}(\chi_B \mid \mathcal{A})) \tag{12}$$

In the notation of conditional probability $\mathscr{P}(\cdot \mid \mathcal{A})$, the $(\mathcal{A}, q_{\mathcal{A}})$-conditional probability (3.2) can also be written as

$$q(B) = \int_X \mathscr{P}(B \mid \mathcal{A}) dq_{\mathcal{A}} \tag{13}$$

Note that both the state $\psi_{\mathcal{A}}$ and its extension ψ in the definition of conditional probability (equations (11) and (12)) are $\|\cdot\|_1$-continuous linear functionals (on $\mathcal{L}^1(X, \mathcal{A}, p_{\mathcal{A}})$ and $\mathcal{L}^1(X, \mathcal{S}, p)$, respectively). Thus, by Remark 2.1, the conditional probability measure, i.e., the restriction of ψ to (the characteristic functions of the) σ-algebra $\mathcal{S}$, is absolutely continuous with respect to p. Thus the concept of Bayesian conditional probability formulated in terms of conditional expectations shares with simple Bayes rule the feature that a conditional probability is always absolutely continuous with respect to the prior probability measure. In contrast to conditional probability defined by Bayes rule, a conditional probability given by Definitions 3.1 and 3.2 can be faithful however (see Example 7.2). Also by Remark 2.1, if $f, g \in \mathcal{L}^1(X, \mathcal{S}, p)$ and $f = g$, p-almost everywhere, then their conditional expected values are equal: $\psi(f) = \psi(g)$. Thus conditionalization and conditional probabilities can be treated, without loss of content, on the factor space $L^1(X, \mathcal{S}, p)$ rather than on $\mathcal{L}^1(X, \mathcal{S}, p)$. This is what we will do from §4 on.

We now show that both Jeffrey conditionalization and elementary conditionalization via the Bayes rule are particular cases of the conditional probability defined via conditional expectations in the manner given by Definition 3.2. To see this recall first a well-known fact from the theory of conditional expectations:

PROPOSITION 3.3 ([1, p. 446], [2, p. 340]). *Let* $(X, \mathcal{S}, p)$ *be a probability space. If the* σ*-subalgebra* $\mathcal{A}$ *of* $\mathcal{S}$ *is generated by a countably infinite partition* $\{A_i\}_{i\in\mathbb{N}}$ *such that* $p(A_i) \neq 0$ $(i = 1, \ldots)$*, then the conditional expectation* (5) *can be given explicitly on the characteristic functions of* $\mathcal{L}^1(X, \mathcal{S}, p)$ *as*

$$\mathscr{E}(\chi_B \mid \mathcal{A}) = \sum_i \frac{p(B \cap A_i)}{p(A_i)} \chi_{A_i} \qquad \forall B \in \mathcal{S} \tag{14}$$

In particular, if $\mathcal{S}$ *has a finite number of elements, then all conditional expectations are of the above form* (*with a finite summation*).

It follows that if state $\psi_{\mathcal{A}}$ is given on $\mathcal{A}$ by fixing its values $\psi_{\mathcal{A}}(\chi_{A_i}) = q_{\mathcal{A}}(A_i)$ on the generating sets A_i, then the $(\mathcal{A}, \psi_{\mathcal{A}})$-conditional probability $q(B)$ of $B \in \mathcal{S}$, $B \notin \mathcal{A}$ that Definition 3.2 specifies is

$$\begin{aligned} q(B) \doteq \psi(\chi_B) &= \psi_{\mathcal{A}}(\mathscr{E}(\chi_B \mid \mathcal{A})) & (15)\\ &= \psi_{\mathcal{A}}\Big(\sum_i \frac{p(B \cap A_i)}{p(A_i)} \chi_{A_i}\Big) & (16)\\ &= \sum_i \frac{p(B \cap A_i)}{p(A_i)} \psi_{\mathcal{A}}(\chi_{A_i}) & (17)\\ &= \sum_i \frac{p(B \cap A_i)}{p(A_i)} q_{\mathcal{A}}(A_i) & (18) \end{aligned}$$

For a finite partition (15)–(18) is the Jeffrey conditional rule [26]. Note that if $\mathcal{A}$ is generated by a countably infinite set $\{A_i\}_{i\in\mathbb{N}}$ of pairwise orthogonal elements from $\mathcal{S}$ but $p(A_i) = 0$ for some A_i then (14) still yields the conditional expectation with the modification that the undefined $\frac{p(B\cap A_i)}{p(A_i)}$ is replaced by any number—this is the phenomenon of the conditional expectation being defined up to a probability zero set (Proposition 2.4). This freedom also is mentioned in connection with Jeffrey conditionalization, see e.g., [44].

Simple Bayesian conditioning is a special case of Jeffrey conditioning: If the Boolean algebra $\mathcal{A}$ in Proposition 3.3 is generated by two nontrivial elements $A, A^\perp$ and we take $\psi_{\mathcal{A}}$ to be the special state on the Boolean algebra $\mathcal{A}$ that takes the values $\psi_{\mathcal{A}}(\chi_A) = q_{\mathcal{A}}(A) = 1$ and $\psi_{\mathcal{A}}(\chi_{A^\perp}) = q_{\mathcal{A}}(A^\perp) = 0$, then the Jeffrey conditionalization rule (15)–(18) reduces to Bayes' rule:

$$\begin{aligned} q(B) &= \frac{p(B \cap A)}{p(A)} \psi_{\mathcal{A}}(\chi_A) + \frac{p(B \cap A^\perp)}{p(A)} \psi_{\mathcal{A}}(\chi_{A^\perp}) & (19)\\ &= \frac{p(B \cap A)}{p(A)} & (20) \end{aligned}$$

REMARK 3.4. *In light of recovering Bayes' rule this way as a special case of conditioning via conditional expectation it becomes visible that the simple Bayes' rule* (19)–(20) *is slightly deceptive: Bayes' rule gives the impression that it is the probability measure* p *that gets conditionalized by* A*. But in fact it is the specific probability measure* $q_{\mathcal{A}}$ *having the particular values* $q_{\mathcal{A}}(A) = 1$ *on* A *and* $q_{\mathcal{A}}(A^\perp) = 0$ *on* $A^\perp$ *that "gets conditionalized"* (*i.e., extended from the Boolean algebra* $\mathcal{A}$ *generated by* A *and* $A^\perp$) *to a probability measure on the whole* σ*-algebra* $\mathcal{S}$*—the role of the probability measure* p *is to serve as the background measure with respect to which the extension takes place. Thus Bayes' rule conceals somewhat the true logical structure of conditionalization, which is the following*:

(i) *The measure $q_{\mathcal{A}}$ on the subalgebra $\mathcal{A}$ represents the conditioning conditions.*

(ii) *The extension q of $q_{\mathcal{A}}$ to the whole algebra $\mathcal{S}$ yields the $(\mathcal{A}, q_{\mathcal{A}})$-conditional probability on condition that the values of q are prescribed on $\mathcal{A}$.*

(iii) *p is the* fixed *background probability measure with respect to which the conditioned values are obtained from $q_{\mathcal{A}}$ via Bayesian statistical inference.*

It must be emphasized that conditionalizing using the theory of conditional expectations in the spirit of Definitions 3.1 and 3.2 is much more general than the Jeffrey conditionalization: a general $\mathcal{A}$ is not generated by a countable partition, and in such cases the $\mathcal{A}$-conditional expectation cannot be of the form (14). A classic example is the Borel-Kolmogorov Paradox situation, which can be defused using the technique of conditional expectations [1, p. 441], [29, p. 50–51], [9, 15, 39] (although the issue remains controversial [17], [25, p. 470], [34]). The $\mathcal{A}$-conditional expectation cannot even always be given explicitly, its existence is the corollary of the Radon–Nikodym theorem, which is a nonconstructive, pure existence theorem.

It should be noted that Jeffrey mentioned the issue of generalization of his rule of conditionalizing in order to include the "continuous case" [26, sec. 11.8]. It was also clear to him that "To discuss the matter more rigorously and generally, it is necessary to use the notion of integration over abstract spaces..." [26, p. 177]. But he did not seem to have worked out the general case systematically. Nor did he refer to the theory of conditional expectations, which is precisely the theory developed by Kolmogorov to cover the general case. Expositions of the mathematical theory of conditional expectations, nowadays a standard topic in probability theory, also do not refer to Jeffrey conditionalization. It is not clear to us why the connection has not been made, although, as we have seen, the connection is straightforward.

§4. The Bayes accessibility relation in terms of conditional expectations.

DEFINITION 4.1. *If ϕ is a state in $L^1(X, \mathcal{S}, p)^\sharp$ then we say that ϕ is Bayes accessible for the Bayesian Agent if there exists a* proper *σ-subalgebra $\mathcal{A}$ of $\mathcal{S}$ and a state $\psi_{\mathcal{A}}$ in $L^1(X, \mathcal{A}, p_{\mathcal{A}})^\sharp$ such that conditionalizing $\psi_{\mathcal{A}}$ using the conditional expectation*

$$\mathbb{E}(\cdot \mid \mathcal{A})\colon L^1(X, \mathcal{S}, p) \to L^1(X, \mathcal{A}, p_{\mathcal{A}}) \tag{21}$$

we obtain ϕ, i.e., if we have

$$\phi(f) = \psi_{\mathcal{A}}(\mathbb{E}(f \mid \mathcal{A})) \qquad \textit{for all } f \in L^1(X, \mathcal{S}, p). \tag{22}$$

The Bayesian interpretation of Bayes accessibility of ϕ is straightforward: The probability measure p represents the background knowledge of the Bayesian Agent ($p_{\mathcal{A}}$ is the restriction of p to the subalgebra $\mathcal{A}$). Suppose state ϕ describes the expectation values of random variables determined by some probability distribution that is given objectively. The Agent might wish to know what this state ϕ is. If ϕ is Bayes accessible for the Agent, then there exists a set of propositions represented by a proper σ-subalgebra $\mathcal{A}$ of $\mathcal{S}$ such that from the evidence given by state $\psi_{\mathcal{A}}$ on the proper subspace $L^1(X, \mathcal{A}, p_{\mathcal{A}})$ determined by the proper subalgebra $\mathcal{A}$ and by the probability representing the Agent's background measure, the Agent can infer and thus learn the values of ϕ by Bayesian statistical inference, i.e., by Bayesian conditionalizing $\psi_{\mathcal{A}}$ using conditional expectations as the conditioning device.

The qualification in Definition 4.1 that $\mathcal{A}$ has to be a *proper* σ-subalgebra is important because every state can be obtained from itself trivially by using the identity map as conditional expectation. To investigate the features of Bayesian learning it is useful however to drop the qualification in Definition 4.1 that $\mathcal{A}$ has to be a *proper* σ-subalgebra and define a general Bayes accessibility relation as follows:

DEFINITION 4.2. *If ϕ and ψ are states in $L^1(X, \mathcal{S}, p)^\sharp$ then we say that ϕ is Bayes accessible from ψ (which we denote by $\psi \overset{\mathbb{E}}{\rightsquigarrow} \phi$), if there exists a (not necessarily proper) σ-subalgebra $\mathcal{A}$ of $\mathcal{S}$ such that conditionalizing ψ using the conditional expectation*

$$\mathbb{E}(\cdot \mid \mathcal{A})\colon L^1(X, \mathcal{S}, p) \to L^1(X, \mathcal{A}, p_{\mathcal{A}}) \tag{23}$$

we obtain ϕ; i.e., if we have

$$\phi(f) = \psi(\mathbb{E}(f \mid \mathcal{A})) \qquad \textit{for all } f \in L^1(X, \mathcal{S}, p) \tag{24}$$

The relation of Definitions 4.1 and 4.2 is straightforward: Since the range of the conditional expectation $\mathbb{E}(\cdot \mid \mathcal{A})$ is the subspace $L^1(X, \mathcal{A}, p_{\mathcal{A}})$ of $L^1(X, \mathcal{S}, p)$, from the perspective of Bayes accessibility of ϕ from ψ only the values of ψ on $L^1(X, \mathcal{A}, p_{\mathcal{A}})$ matter. Thus, if there exists a state $\psi_{\mathcal{A}}$ on the *proper subspace* $L^1(X, \mathcal{A}, p_{\mathcal{A}})$ from which ϕ can be obtained by conditioning using the conditional expectation $\mathbb{E}(\cdot \mid \mathcal{A})$ (and hence ϕ is Bayes accessible for the Agent), then ϕ can be Bayes accessed from any extension of $\psi_{\mathcal{A}}$ to a state ψ on $L^1(X, \mathcal{S}, p)$. Conversely, if for a state ϕ there exists a state ψ in $L^1(X, \mathcal{S}, p)^\sharp$ such that $\psi \overset{\mathbb{E}}{\rightsquigarrow} \phi$ and $\psi \neq \phi$, then ϕ is Bayes accessible for the Bayesian Agent in the sense of Definition 4.1 because the requirement $\psi \neq \phi$ entails that the conditional expectation yielding ϕ from ψ cannot be the identity map and therefore its range is a proper linear subspace $L^1(X, \mathcal{A}, p_{\mathcal{A}})$ determined by a proper subalgebra $\mathcal{A}$. Thus "ϕ is Bayes accessible for the Bayesian Agent" is equivalent to "ϕ is Bayes accessible from *some* $\psi \neq \phi$".

The Bayes accessibility (Definition 4.2) defines a two-place relation in the state space $L^1(X, \mathcal{S}, p)^\sharp$, a subset of the dual space $L^1(X, \mathcal{S}, p)^*$ of the space of integrable random variables $L^1(X, \mathcal{S}, p)$. This Bayes accessibility relation is given by the dual $\mathbb{E}(\cdot \mid \mathcal{A})^*$ of conditional expectations $\mathbb{E}(\cdot \mid \mathcal{A})$, where the dual $\mathbb{E}(\cdot \mid \mathcal{A})^*$ of $\mathbb{E}(\cdot \mid \mathcal{A})$ is defined by

$$L^1(X, \mathcal{S}, p)^\sharp \ni \phi \mapsto \mathbb{E}(\cdot \mid \mathcal{A})^*\phi \doteq \phi \circ \mathbb{E}(\cdot \mid \mathcal{A}) \in L^1(X, \mathcal{S}, p)^\sharp \tag{25}$$

Using the Bayes accessibility relation one can define another two-place relation, which relates elements in the set of all probability measures on $\mathcal{S}$:

DEFINITION 4.3. *Given the measurable space $(X, \mathcal{S})$, we say that two probability measures p and q on σ-algebra $\mathcal{S}$ stand in the prior-posterior relation (equivalently: that (p, q) is a prior-posterior pair) if the linear map ϕ_q defined by*

$$L^1(X, \mathcal{S}, p) \ni f \mapsto \phi_q(f) \doteq \int f\, dq \tag{26}$$

belongs to $L^1(X, \mathcal{S}, p)^\sharp$ and ϕ_q is Bayes accessible from some state ψ in $L^1(X, \mathcal{S}, p)^\sharp$ in the sense of Definition 4.2. The prior-posterior pair (p, q) is called nontrivial *if ϕ_q is Bayes accessible from some state ψ different from ϕ_q, i.e., if equation* (24) *in Definition 4.2 is satisfied with a nontrivial proper subalgebra $\mathcal{A}$ of $\mathcal{S}$.*

The prior-posterior relation connects the inferred probability not to the evidence from which it was inferred but to the background probability p that determines the conditional expectation used in the inference. Hence, if for a given q there is a p such that (p, q) is a *nontrivial* prior-posterior pair, this means that there exists a Bayesian Agent (namely the one having p as background probability) for whom q is Bayes accessible in the sense of Definition 4.1. That is to say, this Bayesian Agent can learn q from some evidence via conditionalization using the conditional expectation determined by p. In the prior-posterior relation the background measure is viewed as a variable. In contrast, in the Bayes accessibility relation the background measure is fixed, and the Bayes accessibility relation reflects what a Bayesian Agent can learn from evidence on the basis of a given, fixed background probability. These two relations thus express different aspects of Bayesian learning.

The prior-posterior relation is fairly easy to characterize: (p, q) is a prior-posterior pair if and only if q is absolutely continuous with respect to p. The necessity of absolute continuity is part of the definition of the prior-posterior relation because a probability measure defined by a state in $L^1(X, \mathcal{S}, p)^\sharp$ is absolutely continuous with respect to p (cf. Remark 2.1). To see that absolute continuity of q with respect to p is sufficient for (p, q) to be a (not necessarily nontrivial) prior-posterior pair, note that if q is absolutely continuous with respect to p, then by the Radon–Nikodym theorem $\phi_q \in L^1(X, \mathcal{S}, p)^\sharp$ and ϕ_q is then Bayes accessible from the state $\psi = \phi_q$ (i.e., from itself) because:

$$\phi_q(f) = \phi_q\left(\mathbb{E}_p(f \mid \mathcal{S})\right) = \phi_q\left(Id(f)\right) \tag{27}$$

where Id is the identity conditional expectation. Thus the prior-posterior relation has exactly the same properties the absolute continuity relation has: it is reflexive, transitive, and it is not symmetric and not antisymmetric.

Since q has to be absolutely continuous with respect to p for (p, q) to be a prior-posterior pair, it is not true that any q can be Bayes learned on the basis of any prior—this is a well-known phenomenon already present in the simplest case of statistical inference based on Bayes' rule: it is impossible to raise zero probability to nonzero by conditionalization. One can show however that for any "sufficiently nontrivial" q there is a prior such that (p, q) is a *nontrivial* prior-posterior pair; i.e., that ψ_q is Bayes accessible for the Agent having p as his background probability. The precise condition of "sufficiently nontrivial" is spelled out in the hypotheses of the following proposition:

PROPOSITION 4.4. *Let $(X, \mathcal{S})$ be a measurable space and q be a probability measure on $\mathcal{S}$ such that*

(A) *There is an event $A \in \mathcal{S}$ with $q(A) \neq 0$ and $q(A) \neq 1$.*
(B) *Both A and $A^\perp$ can be split into two subsets each having non-zero q-probability.*

Then there exists a probability measure p on $\mathcal{S}$ such that (p, q) is a non-trivial *prior-posterior pair.*

Proof. For simplicity let us assume that the support of q is X; i.e., that the density function (Radon–Nikodym derivative) $\frac{dq}{dq}$ is the constant function with value 1 on the entire X. We prove first that there is a function $h : X \to \mathbb{R}$ which is constant on A and that there is a measure p on $\mathcal{S}$ such that

$$q(H) = \int_X h\chi_H \, dp \qquad \text{for all } H \in \mathcal{S}. \tag{28}$$

Of course, in this case h is the Radon–Nikodym derivative $\frac{dq}{dp}$. Let $f : X \to \mathbb{R}$ be a function such that $f(x) = 1$ for $x \in A$, $f(x) > 0$ for all $x \in X$, $\int_X f \, dq = 1$ and such that f is not the constant function with value 1 on the entire X. Such an f exists because by stipulation (B) event $A^\perp$ can be decomposed into non-q-measure zero events $B_1 \cup B_2$ and thus f can be chosen so that it is constant on B_1 and B_2 with different constants.

Define the measure p on $\mathcal{S}$ by

$$p(H) = \int_X f\chi_H \, dq \qquad \text{for all } H \in \mathcal{S}. \tag{29}$$

This p will be the sought-after prior measure. As $p(H) = \int \chi_H \; dp = \int f \chi_H \; dq$, it follows that f is the Radon–Nikodym derivative $\frac{dp}{dq}$. Since $f(x) > 0$ and $\frac{dq}{dq} > 0$ everywhere, the probabilities p and q are mutually absolutely continuous, therefore [18, sec. 32, p. 136] the density functions $\frac{dp}{dq}$ and $\frac{dq}{dp}$ satisfy the equation

$$\frac{dp}{dq} = \frac{1}{\frac{dq}{dp}}$$

This implies

$$q(H) = \int_X \frac{dq}{dp} \chi_H \; dp = \int_X \frac{1}{f} \chi_H \; dp. \tag{30}$$

So we take $h = \frac{1}{f}$. Note that h is constant on A.

Let $\mathcal{A}$ be the σ-subalgebra of $\mathcal{S}$ generated by A and all the elements $\{x\}$ with $x \in A^\perp$. A simple argument shows (cf. equation (6)) that for any measure p we have

$$\mathbb{E}_p(v \mid \mathcal{A})(x) = \begin{cases} \frac{1}{p(A)} \int_A v \; dp & \text{if } x \in A \\ v(x) & \text{otherwise} \end{cases} \qquad \forall v \in L^1(X, \mathcal{S}, p).$$

By stipulations (A) and (B) the σ-subalgebra $\mathcal{A}$ is a proper subalgebra of $\mathcal{S}$. We need to show now that there is a state $\psi \in L^1(X, \mathcal{S}, p)^\sharp$ such that

$$\phi_q(v) = \psi\left(\mathbb{E}_p(v \mid \mathcal{A})\right) \qquad \forall v \in L^1(X, \mathcal{S}, p) \tag{31}$$

Any $\psi \in L^1(X, \mathcal{S}, p)^\sharp$ can be represented by a density function g with respect to p: $\psi(v) = \int vg \; dp$ for all $v \in L^1(X, \mathcal{S}, p)$. Therefore equation (31) is equivalent to

$$\phi_q(v) = \int_X v \; dq = \int_X vh \; dp = \int_X .g\mathbb{E}_p(v \mid \mathcal{A}) \; dp. \tag{32}$$

Let us define g by

$$g(x) = \begin{cases} \int_A h \; dp & \text{for } x \in A \\ h(x) & \text{otherwise} \end{cases}$$

As h is constant on A we have

$$\begin{aligned} \int_A vh \; dp &= \int_A h \; dp \cdot \int_A v \; dp & (33)\\ &= \int_A g \cdot \mathbb{E}_p(v \mid \mathcal{A}) \; dp & (34)\\ &= \int_A \left(\int_A h \; dp \cdot \frac{\int_A v \; dp}{p(A)} \right) dp & (35) \end{aligned}$$

And for $A^\perp$ we have

$$\int_{X-A} vh \; dp = \int_{X-A} g \cdot \mathbb{E}_p(v \mid \mathcal{A}). \tag{36}$$

Therefore by letting $\psi(v) = \int_X v \cdot g \; dp$ we obtain equation (31) and this completes the proof. □

Stipulations (A) and (B) in Proposition 4.4 can be satisfied easily; in fact, in standard applications they automatically hold: For a finite or countable measurable space $(X, \mathcal{S})$ it is enough to suppose that the support of q contains at least 4 elements. Conditions (A) and (B) also hold for all probability measures that are absolutely continuous with respect to the Lebesgue measure. There exist however "exotic" probability spaces for which conditions (A) and (B) do not hold. (See the example after Proposition 6.4.) A careful examination of the proof of Proposition 4.4 reveals that (p, q) being a nontrivial prior-posterior pair is intimately related to the Radon–Nikodym derivative $\frac{dq}{dp}$ being noninjective. Indeed, this idea will be made precise in Proposition 6.2.

Every Bayesian Agent comes equipped with some background measure. It is an important epistemological question then: What are the epistemological ramifications of the Bayesian Agent having a particular background measure. The answer to this question is contained in the features of the Bayes accessibility relation. This is what we study in the following sections.

§5. Antisymmetry of the Bayes accessibility relation and failure of strong Bayes connectedness of state spaces. We have already noted in the previous section that every state is Bayes accessible from itself: Taking the identity map on $L^1(X, \mathcal{S}, p)$ as the conditional expectation $\mathbb{E}$ one can obtain any state ϕ as its own conditioned state. This means that the Bayes accessibility relation $\overset{\mathbb{E}}{\rightsquigarrow}$ is reflexive.

PROPOSITION 5.1. *The relation $\overset{\mathbb{E}}{\rightsquigarrow}$ is antisymmetric.*

Proof. Assume $\psi \overset{\mathbb{E}}{\rightsquigarrow} \phi$ and $\phi \overset{\mathbb{E}}{\rightsquigarrow} \psi$. Then there exist σ-subalgebras $\mathcal{S}_1, \mathcal{S}_2$ of $\mathcal{S}$ and conditional expectations

$$\mathbb{E}(\cdot \mid \mathcal{S}_1) : L^1(X, \mathcal{S}, p) \to L^1(X, \mathcal{S}_1, p_1) \tag{37}$$

$$\mathbb{E}(\cdot \mid \mathcal{S}_2) : L^1(X, \mathcal{S}, p) \to L^1(X, \mathcal{S}_2, p_2). \tag{38}$$

Such that

$$\phi(f) = \psi(\mathbb{E}(f \mid \mathcal{S}_1)) \qquad \forall f \in L^1(X, \mathcal{S}, p) \tag{39}$$

$$\psi(f) = \phi(\mathbb{E}(f \mid \mathcal{S}_2)) \qquad \forall f \in L^1(X, \mathcal{S}, p). \tag{40}$$

Let $\mathbb{E}_1$ and $\mathbb{E}_2$ denote the orthogonal projections on $L^2(X, \mathcal{S}, p)$ corresponding to the conditional expectations $\mathbb{E}(\cdot \mid \mathcal{S}_1)$ and $\mathbb{E}(\cdot \mid \mathcal{S}_2)$. Equations (39)–(40) entail then

$$\phi(f) = \psi(\mathbb{E}_1 f) \qquad \forall f \in L^2(X, \mathcal{S}, p) \tag{41}$$

$$\psi(f) = \phi(\mathbb{E}_2 f) \qquad \forall f \in L^2(X, \mathcal{S}, p) \tag{42}$$

Equations (41)–(42) entail

$$\phi(f) = \psi(\mathbb{E}_1\mathbb{E}_2\mathbb{E}_1 f) \qquad \forall f \in L^2(X, \mathcal{S}, p) \tag{43}$$

$$\psi(f) = \phi(\mathbb{E}_2\mathbb{E}_1\mathbb{E}_2 f) \qquad \forall f \in L^2(X, \mathcal{S}, p) \tag{44}$$

and equations (43)–(44) entail that for all $n \in \mathbb{N}$ we have

$$\phi(f) = \psi([\mathbb{E}_1\mathbb{E}_2\mathbb{E}_1]^n f) \qquad \forall f \in L^2(X, \mathcal{S}, p) \tag{45}$$

$$\psi(f) = \psi([\mathbb{E}_2\mathbb{E}_1\mathbb{E}_2]^n f) \qquad \forall f \in L^2(X, \mathcal{S}, p) \tag{46}$$

Since ϕ and ψ are assumed to be $\| \cdot \|_1$-continuous (45)–(46) entail:

$$\phi(f) = \psi(\overset{1}{\lim_{n\to\infty}}([\mathbb{E}_1\mathbb{E}_2\mathbb{E}_1]^n f)) \qquad \forall f \in L^2(X, \mathcal{S}, p) \tag{47}$$

$$\psi(f) = \phi(\overset{1}{\lim_{n\to\infty}}([\mathbb{E}_2\mathbb{E}_1\mathbb{E}_2]^n f)) \qquad \forall f \in L^2(X, \mathcal{S}, p) \tag{48}$$

where $\overset{1}{\lim}$ denotes the limit in the $\| \cdot \|_1$ norm.

Since $\mathbb{E}_1$ and $\mathbb{E}_2$ are projections on the Hilbert space $\mathcal{H} = L^2(X, \mathcal{S}, p)$, the limits of the operator sequences $[\mathbb{E}_1\mathbb{E}_2\mathbb{E}_1]^n$ and $[\mathbb{E}_2\mathbb{E}_1\mathbb{E}_2]^n$ exist in the sense of the strong operator topology in the set of all bounded operators $\mathcal{B}(\mathcal{H})$ on $\mathcal{H}$, the limits are the same, and the limit is an element in the lattice $\mathcal{P}(\mathcal{H})$ of all projections on $\mathcal{H}$: it is the greatest lower bound $\mathbb{E}_2 \wedge \mathbb{E}_1$ of the the projections $\mathbb{E}_1$ and $\mathbb{E}_2$ with respect to the standard ordering $\leq$ of projections in $\mathcal{P}(\mathcal{H})$ (Proposition 4.13 in [38]). So for all $f \in L^2(X, \mathcal{S}, p)$ we have

$$\overset{2}{\lim_{n\to\infty}} [\mathbb{E}_1\mathbb{E}_2\mathbb{E}_1]^n f = \overset{2}{\lim_{n\to\infty}} [\mathbb{E}_2\mathbb{E}_1\mathbb{E}_2]^n f \tag{49}$$

$$= (\mathbb{E}_2 \wedge \mathbb{E}_1) f \tag{50}$$

where $\overset{2}{\lim}$ denotes the limit in the $\|\cdot\|_2$ norm.

Since p is a bounded measure, by Jensen's inequality one has $\|f\|_1 \leq \|f\|_2$; hence the limit of the sequences $[\mathbb{E}_1\mathbb{E}_2\mathbb{E}_1]^n f$ and $[\mathbb{E}_2\mathbb{E}_1\mathbb{E}_2]^n f$ also exists in the $\|\cdot\|_1$ norm, and so for all $f \in L^2(X, \mathcal{S}, p)$ we have

$$\overset{1}{\lim_{n\to\infty}} [\mathbb{E}_1\mathbb{E}_2\mathbb{E}_1]^n f = \overset{1}{\lim_{n\to\infty}} [\mathbb{E}_2\mathbb{E}_1\mathbb{E}_2]^n f \tag{51}$$

$$= (\mathbb{E}_2 \wedge \mathbb{E}_1) f \tag{52}$$

Equations (47)–(48) together with (51)–(52) entail:

$$\phi(f) = \psi((\mathbb{E}_2 \wedge \mathbb{E}_1) f) \qquad \forall f \in L^2(X, \mathcal{S}, p) \tag{53}$$

$$\psi(f) = \phi((\mathbb{E}_1 \wedge \mathbb{E}_2) f) \qquad \forall f \in L^2(X, \mathcal{S}, p) \tag{54}$$

Since $\mathbb{E}_i \geq (\mathbb{E}_2 \wedge \mathbb{E}_1)$ $(i = 1, 2)$, we have

$$\mathbb{E}_i(\mathbb{E}_2 \wedge \mathbb{E}_1) = (\mathbb{E}_2 \wedge \mathbb{E}_1) \qquad i = 1, 2 \tag{55}$$

equations (41)–(42) and (53)–(54) entail that for all $f \in L^2(X, \mathcal{S}, p)$ we have

$$\phi(f) = \psi((\mathbb{E}_2 \wedge \mathbb{E}_1) f) \tag{56}$$

$$= \phi(\mathbb{E}_2(\mathbb{E}_2 \wedge \mathbb{E}_1) f) \tag{57}$$

$$= \phi((\mathbb{E}_2 \wedge \mathbb{E}_1) f) \tag{58}$$

$$= \psi(f). \tag{59}$$

Thus ϕ is equal to ψ on $L^2(X, \mathcal{S}, p)$. Since the $L^2(X, \mathcal{S}, p)$ is $\|\cdot\|_1$-dense in $L^1(X, \mathcal{S}, p)$ and ϕ and ψ are $\|\cdot\|_1$-continuous, ϕ and ψ are equal on $L^1(X, \mathcal{S}, p)$. □

Antisymmetry of the Bayes accessibility relation $\overset{\mathbb{E}}{\rightsquigarrow}$ entails that state spaces are *not strongly Bayes connected* in general: it is not true that any state ϕ in $L^1(X, \mathcal{S}, p)^\sharp$ is Bayes accessible from any other ψ in $L^1(X, \mathcal{S}, p)^\sharp$. If strong Bayes connectedness were a feature of a state space then every state could be learned by the Agent by Bayesian upgrading from every other in the same state space: Given any two states ϕ and ψ there would always exist a set of propositions (depending on ϕ of course) such that knowing the values of ψ on elements of that set, the Agent could infer all values of ϕ by conditionalizing ψ (with respect to the fixed background probability measure) on that set of propositions. But antisymmetry of $\overset{\mathbb{E}}{\rightsquigarrow}$ entails that the only probability space that is strongly Bayes connected is the trivial one with $\emptyset$ and X forming $\mathcal{S}$. Thus in nontrivial state spaces Bayesian learning has a certain directedness: if ϕ can be Bayes-learned from ψ, then ψ cannot be

Bayes-learned from ϕ. What can be Bayes-learned from some evidence, cannot serve as evidence to Bayes-learn the evidence itself.

§6. Are state spaces weakly Bayes connected? Lack of strong Bayes connectedness of state spaces leads to the following definition:

DEFINITION 6.1. *A state space $L^1(X, \mathcal{S}, p)^\sharp$ is called* weakly *Bayes connected if for every state ϕ in $L^1(X, \mathcal{S}, p)^\sharp$ there exists a state ψ in $L^1(X, \mathcal{S}, p)^\sharp$ such that $\psi \neq \phi$ and ϕ is Bayes accessible from ψ.*

Are state spaces weakly Bayes connected? There is no general "yes" or "no" answer to this question. We will see that some state spaces are, some others are not weakly Bayes connected. To decide whether a state space is weakly Bayes connected is not a trivial task. For instance we do not know whether a "small enough" state space is weakly Bayes connected (cf. Problem 6.6) (we conjecture that it is not). We show here failure of weak Bayes connectedness of state spaces of typical probability theories, and give an example of a weakly Bayes connected state space. To do this, we have to separate the analysis of weak Bayes connectedness into two parts: considering first the space $L^2(X, \mathcal{S}, p)$ and then $L^1(X, \mathcal{S}, p)$. We start with the L^2-theory of Bayes connectedness of state spaces.

Recall that two probability spaces $(X_1, \mathcal{S}_1, p_1)$ and $(X_2, \mathcal{S}_2, p_2)$ are *isomorphic* if there is an invertible map $f : X_1 \to X_2$ such that both f and f^{-1} are measurable, measure preserving maps. A closely related notion is isomorphism modulo zero: $(X_1, \mathcal{S}_1, p_1)$ and $(X_2, \mathcal{S}_2, p_2)$ are *isomorphic modulo* 0 if there exist sets $A_1 \subseteq X_1$ and $A_2 \subseteq X_2$ with $p_1(A_1) = 0 = p_2(A_2)$ such that the probability spaces $(X'_1, \mathcal{S}'_1, p'_1)$ and $(X'_2, \mathcal{S}'_2, p'_2)$ are isomorphic, where $X'_1 = X_1 \smallsetminus A_1$ and $X'_2 = X_2 \smallsetminus A_2$, $\mathcal{S}'_1$, and $\mathcal{S}'_2$ are the natural restrictions of the σ-algebras $\mathcal{S}_1$ and $\mathcal{S}_2$ obtained by removing the sets A_1 and A_2, and where p'_1 and p'_2 are the restrictions of p_1 and p_2 to $\mathcal{S}'_1$, and $\mathcal{S}'_2$. If $(X_1, \mathcal{S}_1, p_1)$ and $(X_2, \mathcal{S}_2, p_2)$ are isomorphic modulo 0, then $L^s(X_1, \mathcal{S}_1, p_1)$ and $L^s(X_2, \mathcal{S}_2, p_2)$ are isometrically isomorphic spaces (because in L^s spaces functions differing on null sets are identified).

Let $(X, \mathcal{S}, p)$ be a probability space and $\mathcal{A}$ be a σ-subalgebra of $\mathcal{S}$. We say that $\mathcal{A}$ and $\mathcal{S}$ are equal modulo 0 if $(X, \mathcal{S}, p)$ and $(X, \mathcal{A}, p)$ are isomorphic modulo 0. $\mathcal{A}$ and $\mathcal{S}$ being not equal modulo zero means that there is a set $A \in \mathcal{S} \setminus \mathcal{A}$ with $p(A) \neq 0$. Note that $L^s(X, \mathcal{A}, p)$ is always a closed subspace of $L^s(X, \mathcal{S}, p)$ but equality of $\mathcal{A}$ and $\mathcal{S}$ modulo zero implies $L^s(X, \mathcal{S}, p) = L^s(X, \mathcal{A}, p)$.

The next proposition formulates a condition that is equivalent to the weak Bayes connectedness of L^2 state spaces. Throughout $\mathcal{L}$ denotes the Lebesgue σ-algebra over the reals.

PROPOSITION 6.2. *$L^2(X, \mathcal{S}, p)^\sharp$ is weakly Bayes connected if and only if there exists no function $f \in L^2(X, \mathcal{S}, p)$ which is positive $f > 0$, normalized $\|f\|_2 = 1$, and such that $\mathcal{S}$ and $f^{-1}[\mathcal{L}]$ are equal modulo 0, where $f^{-1}[\mathcal{L}] = \left\{f^{-1}(A) : A \in \mathcal{L}\right\}$ with f^{-1} being the inverse image function of f.*

Proof. By Riesz's representation theorem ([1, p. 244]) for each state $\phi \in L^2(X, \mathcal{S}, p)^\sharp$ there exists a positive, normalized function $f_\phi \in L^2(X, \mathcal{S}, p)$ such that

$$\phi(g) = \langle g, f_\phi \rangle \qquad g \in L^2(X, \mathcal{S}, p), \tag{60}$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product in $L^2(X, \mathcal{S}, p)$. Conversely: every positive, normalized function f in $L^2(X, \mathcal{S}, p)$ defines a state ϕ_f on $L^2(X, \mathcal{S}, p)$ by $\phi_f(g) = \langle g, f \rangle$ for all $g \in L^2(X, \mathcal{S}, p)$.

Let $\phi, \psi \in L^2(X, \mathcal{S}, p)^\sharp$ be two states, f_ϕ and f_ψ be the two functions in $L^2(X, \mathcal{S}, p)$ that represent them in the sense of Riesz' representation theorem. If ϕ is Bayes accessible from ψ, then, by definition of Bayes accessibility, there is a σ-subalgebra $\mathcal{A}$ of $\mathcal{S}$ such that

$$\phi(g) = \psi(\mathbb{E}(g \mid \mathcal{A})) \qquad \text{for all } g \in L^2(X, \mathcal{S}, p) \tag{61}$$

Denoting by $\mathbb{E}_\mathcal{A}$ the operator on $L^2(X, \mathcal{S}, p)$ that represents the conditional expectation $\mathbb{E}(\cdot \mid \mathcal{A})$, and using the Riesz representatives f_ϕ and f_ψ of states ϕ and ψ, equation (61) can be re-written as

$$\langle g, f_\phi \rangle = \langle \mathbb{E}_\mathcal{A} g, f_\psi \rangle \qquad \text{for all } g \in L^2(X, \mathcal{S}, p). \tag{62}$$

Since $\mathbb{E}_\mathcal{A}$ is an orthogonal, selfadjoint projection, equation (62) entails

$$\langle g, f_\phi \rangle = \langle \mathbb{E}_\mathcal{A} g, f_\psi \rangle = \langle g, \mathbb{E}_\mathcal{A} f_\psi \rangle \qquad \text{for all } g \in L^2(X, \mathcal{S}, p). \tag{63}$$

The equation $\langle g, f_\phi \rangle = \langle g, \mathbb{E}_\mathcal{A} f_\psi \rangle$ holds for all $g \in L^2(X, \mathcal{S}, p)$ if and only if $f_\phi = \mathbb{E}_\mathcal{A} f_\psi$. Thus we can conclude that if ϕ is Bayes accessible from some state then f_ϕ is in the range of an orthogonal projection $\mathbb{E}_\mathcal{A}$ representing a conditional expectation. It follows that if the state ϕ is Bayes accessible from a state *different from* ϕ, then its representing vector f_ϕ must belong to a *proper* closed linear subspace of $L^2(X, \mathcal{S}, p)$ that has the form $L^2(X, \mathcal{A}, p)$. Since the smallest closed linear subspace in $L^2(X, \mathcal{S}, p)$ to which f_ϕ belongs is $L^2(X, f_\phi^{-1}[\mathcal{L}], p)$, state ϕ is Bayes accessible from another state only if $f_\phi^{-1}[\mathcal{L}] \subset \mathcal{S}$ is a proper subalgebra which is not equal to $\mathcal{S}$ modulo 0 (for if $f_\phi^{-1}[\mathcal{L}]$ is a subalgebra equal to $\mathcal{S}$ modulo 0, then $L^2(X, f_\phi^{-1}[\mathcal{L}], p)$ is equal to $L^2(X, \mathcal{S}, p)$). Consequently, $L^2(X, \mathcal{S}, p)^\sharp$ is weakly Bayes connected only if there is no positive and normalized function $f \in L^2(X, \mathcal{S}, p)$ such that $\mathcal{S}$ and $f^{-1}[\mathcal{L}]$ are equal modulo zero.

Conversely, suppose there exists no positive and normalized function f such that $\mathcal{S}$ and $f^{-1}[\mathcal{L}]$ are equal modulo zero. Then for every state ϕ the function f_ϕ that represents ϕ in the sense of the Riesz representation theorem, $L^2(X, f_\phi^{-1}[\mathcal{L}], p)$ is a proper closed linear subspace of $L^2(X, \mathcal{S}, p)$. By Proposition 2.3 there exist then a conditional expectation $\mathbb{E}(\cdot \mid f_\phi^{-1}[\mathcal{L}])$ from $L^2(X, \mathcal{S}, p)$ onto $L^2(X, f_\phi^{-1}[\mathcal{L}], p)$ and (since $L^2(X, f_\phi^{-1}[\mathcal{L}], p)$ is a *proper* subspace) also a positive, normalized function $f' \notin L^2(X, f_\phi^{-1}[\mathcal{L}], p)$ such that $f_\phi = \mathbb{E}(f' \mid f_\phi^{-1}[\mathcal{L}])$. This entails $\langle g, f_\phi \rangle = \langle g, \mathbb{E}(f' \mid f_\phi^{-1}[\mathcal{L}]) \rangle$ for all $g \in L^2(X, \mathcal{S}, p)$, which is equivalent to $\phi = \psi \circ \mathbb{E}(\cdot \mid f_\phi^{-1}[\mathcal{L}])$ where ψ is the state in $L^2(X, \mathcal{S}, p)^\sharp$ that is Riesz-represented by function f'. Thus every state in $L^2(X, \mathcal{S}, p)^\sharp$ is obtainable as a conditioned state and so the state space $L^2(X, \mathcal{S}, p)^\sharp$ is weakly Bayes connected. □

LEMMA 6.3. *If there is an injective, positive and normalized function $f \in L^2(X, \mathcal{S}, p)$, then $L^2(X, \mathcal{S}, p)^\sharp$ is* not *weakly Bayes connected.*

Proof. If $f : X \to \mathbb{R}$ is injective, then for all $A \in \mathcal{S}$ we have $f^{-1}(f(A)) = A$. This entails $f^{-1}[\mathcal{L}] = \mathcal{S}$ and the statement follows from Proposition 6.2. □

To state the next proposition we need to recall the notion of a *standard* probability space. Intuitively, a probability space is *standard* if it is "the sum" of continuous and discrete parts, where the continuous part is (measure theoretically) isomorphic to an interval with the Lebesgue (or Borel) measure on it, and the discrete part is (measure theoretically) isomorphic to a measure space with a σ-algebra that is either finite or is generated by a countably infinite set. To give a more precise definition one has to define the sum (disjoint union) of measure spaces: Let $(X_i, \mathcal{S}_i, p_i)$ for $i < n \in \mathbb{N}$ be finitely many measure spaces and suppose for convenience that the X_i's are disjoint sets. Define a σ-algebra $\mathcal{S}$ on $X = \bigcup_i X_i$ as follows: Take a subset $A \subseteq X$ to be in $\mathcal{S}$ if and only if $A \cap X_i$ belongs to $\mathcal{S}_i$ for all i. Then the map $p : \mathcal{S} \to \mathbb{R}$ defined by

$$p(A) \doteq \sum_i p_i(A \cap X_i) \quad \text{for all } A \in \mathcal{S} \tag{64}$$

is a measure and the measure space $(X, \mathcal{S}, p)$ is called the disjoint union of the measure spaces $(X_i, \mathcal{S}_i, p_i)$. (For the elementary properties of a disjoint union of measure spaces we refer to [13, sec. 214K].) A probability space is called standard if it is isomorphic modulo zero to the disjoint union of the Borel or Lebesgue measure spaces of a (possibly empty) interval, and a measure space with a σ-algebra that is either finite or is generated by a countably infinite set (cf. Definition 4.5 in [35]). It is not hard to see that the disjoint union of finitely many standard measure spaces is also standard.

Examples of standard probability spaces include all probability spaces with a finite or countably infinite set of elementary events ("discrete" probability spaces) and the n-dimensional Euclidean spaces $\mathbb{R}^n$ with probability given by a density function with respect to the Lebesgue measure on $\mathbb{R}^n$. Also included are the probability spaces where X is a compact subset E of $\mathbb{R}^n$ and the probability on E is given by a density with respect to the restriction of the Lebesgue measure to E. These probability spaces cover essentially all applications of probability.

PROPOSITION 6.4. *Let $(X, \mathcal{S}, p)$ be a probability space. Then $L^2(X, \mathcal{S}, p)^\sharp$ is* not *weakly Bayes connected in the following* (*i*)–(*iii*) *cases*:

(i) *$(X, \mathcal{S}, p)$ is generated by a countable set of point masses, i.e., X is finite or countably infinite.*

(ii) *$(X, \mathcal{S}, p)$ is isomorphic to an interval with the Borel or Lebesgue measure.*

(iii) *$(X, \mathcal{S}, p)$ is a standard probability space.*

Proof. (i) Suppose X is finite or countably infinite. In this case it is clear that there exists a measurable, injective, positive, and integrable $f : X \to \mathbb{R}$. By re-normalization we can also assume that f is normalized. Then (i) follows from Lemma 6.3. For later purposes we note that such an f can always be assumed to be bounded and hence to belong to $L^\infty(X, \mathcal{S}, p) \cap L^2(X, \mathcal{S}, p)$. (Take for instance $X = \mathbb{N}$ and $f(n) = \frac{1}{n+1}$.)

(ii) Without loss of generality we can assume that $(X, \mathcal{S}, p)$ is the Lebesgue space $([0, 1], \mathcal{L}, \lambda)$. We wish to apply Lemma 6.3 again. It is easy to see that there is an injective, positive function $f : [0, 1] \to \mathbb{R}$ (take, for instance, the identity function $\mathsf{id}_{[0,1]}$ on $[0, 1]$). Clearly f is measurable and belongs to $L^2([0, 1], \mathcal{L}, \lambda)$. To make it normalized, divide it by $\|f\|_2$. For later purposes we note $\mathsf{id}_{[0,1]} \in L^\infty(X, \mathcal{S}, p) \cap L^2(X, \mathcal{S}, p)$.

(iii) In this case $(X, \mathcal{S}, p)$ is isomorphic modulo zero to a disjoint union of a (possibly empty) interval with Lebesgue or Borel measure and a countable (possibly empty) set of point masses. Take the union of the two injective, positive functions obtained from cases (i) and (ii) and normalize it to length 1. Then the result follows again from Lemma 6.3. □

Proposition 6.4 shows that probability spaces are typically not weakly Bayes connected. This leads to the question of whether weakly Bayes connected probability spaces exist at all. We show below that they do by isolating a class of probability spaces which have weakly Bayes connected state spaces. However, the spaces in that class are "very large": Call a probability space $(X, \mathcal{S}, p)$ *significantly large* if it is not isomorphic modulo zero to any space $(X', \mathcal{S}', p')$ with $\mathcal{S}'$ having cardinality less than or equal to the cardinality of the set $\mathcal{L}$ of Lebesgue measurable sets. Significantly large probability spaces exist. Consider for instance the following example. Let X be any uncountable set, and $\mathcal{S}$ be the family of sets $A \subseteq X$ with the property that either A or its complement $X \setminus A$ is countable. Then $\mathcal{S}$ is a σ-algebra of subsets of X, and its cardinality $|\mathcal{S}|$ satisfies $|\mathcal{S}| \geq |X|$. Consider the function $p : \mathcal{S} \to [0, 1]$ defined by $p(A) = 0$ if A is countable and $p(A) = 1$ if A is not countable. Then p is a probability measure on $\mathcal{S}$. If $|X| > 2^{2^{\aleph_0}}$, then $(X, \mathcal{S}, p)$ is significantly large. This is because each p-probability zero set is countable, and removing a countable set does not change the cardinality of X. Recall that $|\mathcal{L}| = 2^{2^{\aleph_0}}$.

The next proposition motivates the definition of significantly large probability spaces.

PROPOSITION 6.5. *If $(X, \mathcal{S}, p)$ is significantly large, then $L^2(X, \mathcal{S}, p)^\sharp$ is weakly Bayes connected.*

Proof. $\mathcal{S}$ cannot be equal modulo zero to $f^{-1}[\mathcal{L}]$ for any $f \in L^2(X, \mathcal{S}, p)$, because in this case $\mathcal{S}$ would be equal modulo zero to an algebra of cardinality $|f^{-1}[\mathcal{L}]| \leq |\mathcal{L}|$. Thus the result follows directly from Proposition 6.2. □

Proposition 6.5 establishes a connection between weak Bayes connectedness of the state space of a probability space and cardinality of the σ-algebra of the propositions over which the Bayesian Agent defines probabilities. This proposition gives a sufficient condition for Bayes connectedness to hold: the σ-algebra of propositions must be larger than the σ-algebra in the set of (real) numbers with respect to which measurability of the random variables is required. It remains open whether this condition also is necessary however. We conjecture that it is.

From the perspective of Bayesian learning, the sufficient condition for weak Bayes connectedness contained in Proposition 6.5 is very demanding: The Bayesian Agent must be able to comprehend a set of elementary (atomic) propositions cardinality of which is way beyond even that of the continuum. Whether one should allow such an extremely strong concept of Bayesian Agent, is questionable. Proposition 6.5 and its proof also indicate in what way the demanding condition could in principle be weakened: One can read Proposition 6.5 as saying that the cardinality of the σ-algebra in the field in which the random variables take their value and with respect to which measurability of the random variables are demanded give a lower bound on the cardinality of the σ-algebra of random events for which weak Bayes connectedness can hold. To put it differently: the coarser the random variables the smaller the minimal size of the σ-algebra of random events that allows in principle for the corresponding probabilistic theory to be weakly Bayes connected. Thus, as long as one considers real valued random variables in the standard interpretation as real valued maps that are required to be Borel (or Lebesgue) measurable, the state spaces of usual probability theories will *not* be weakly Bayes connected.

Propositions 6.4 and 6.5 also lead to the following open problem.

PROBLEM 6.6. *Is there a non-standard probability space $(X, \mathcal{S}, p)$ with cardinality $|\mathcal{S}| = |\mathcal{L}|$ such that its state space $L^2(X, \mathcal{S}, p)^\sharp$ is weakly Bayes connected*?

Next, we turn to the question of weak Bayes connectedness of L^1-state spaces.

PROPOSITION 6.7. *If $(X, \mathcal{S}, p)$ is a standard probability space, then $L^1(X, \mathcal{S}, p)^\sharp$ is* not *weakly Bayes connected.*

Proof. The proof is based on the following idea. Suppose $L^2(X, \mathcal{S}, p)^\sharp$ is not weakly Bayes connected. Then there is a positive, normalized function $f \in L^2(X, \mathcal{S}, p)$ witnessing it: the state ϕ_f is not accessible from any other L^2-state (cf. the proof of Proposition 6.2). Since the dual space of $L^1(X, \mathcal{S}, p)$ is $L^\infty(X, \mathcal{S}, p)$, if f happens to belong to $L^\infty(X, \mathcal{S}, p)$ as well, then f defines a state ϕ in $L^1(X, \mathcal{S}, p)^\sharp$ via

$$\phi(g) = \int fg\,dp \qquad \text{for all } g \in L^1(X, \mathcal{S}, p). \tag{65}$$

We claim that such a ϕ is not Bayes accessible from any other state $\psi \in L^1(X, \mathcal{S}, p)^\sharp$; thus this state will witness $L^1(X, \mathcal{S}, p)^\sharp$ not being weakly Bayes connected.

To see that ϕ is not Bayes accessible recall that L^1-states are L^2-states as well because $\|\cdot\|_1 \leq \|\cdot\|_2$ holds due the fact that p is a bounded measure. Consequently $\|\cdot\|_1$-continuity of a state $\psi \in L^1(X, \mathcal{S}, p)^\sharp$ implies $\|\cdot\|_2$-continuity of ψ. Thus, if ϕ were Bayes accessible from $\psi \in L^1(X, \mathcal{S}, p)^\sharp$, then the same ψ (being an $L^2(X, \mathcal{S}, p)$-state as well) would show that the restriction ϕ_f of ϕ to $L^2(X, \mathcal{S}, p)$ is Bayes accessible; a clear contradiction. Thus all one has to prove is that there is a function $f \in L^2(X, \mathcal{S}, p) \cap L^\infty(X, \mathcal{S}, p)$ witnessing that $L^2(X, \mathcal{S}, p)^\sharp$ is not weakly Bayes connected. But this has essentially been done in the proof of Proposition 6.4. □

Though probability spaces with finite Boolean algebras are standard hence their state spaces not weakly Bayes connected, we include here another proof of violation of weak Bayes connectedness for the finite case. We do this for two reasons: First, because the proof in the finite case shows more explicitly how violation of weak Bayes connectedness occurs. Second, the proof will display an explicit prescription that can be used to obtain a lot of Bayes inaccessible states.

PROPOSITION 6.8. *The state space of $L^1(X, \mathcal{S}, p)$ is not weakly Bayes connected if the cardinality of the σ-algebra $\mathcal{S}$ is finite.*

Proof. Let $(X_n, \mathcal{S}_n, p_n)$ be a probability space with X_n having $n < \infty$ number of elements and with $\mathcal{S}_n$ being the Boolean algebra of the power set of X_n. Let $L^1(X_n, \mathcal{S}_n, p_n)$ be the associated function space. Without loss of generality we may assume that the probability measure p_n is faithful, i.e., $p_n(\{x_i\}) \neq 0$ for every $i = 1, 2, \ldots n$. This is because in the function space $L^1(X_n, \mathcal{S}_n, p_n)$ functions differing on p_n-probability zero sets only are identified, hence if $p_n(\{x_i\}) = 0$ then the characteristic function $\chi_{\{x_i\}}$ of $\{x_i\}$ is the zero element in $L^1(X_n, \mathcal{S}_n, p_n)$. Consequently, $L^1(X_n, \mathcal{S}_n, p_n)$ and $L^1(X_m, \mathcal{S}_m, p_m)$ will be equal, where $(X_m, \mathcal{S}_m, p_m)$ $(m \leq n)$ is obtained from $(X_n, \mathcal{S}_n, p_n)$ by leaving out from X_n the p_n-probability zero events and taking $p_m(\{x_j\}) = p_n(\{x_j\})$ on X_m whenever $p_n(\{x_j\}) \neq 0$. The probability measure p_m is faithful then, and the state space of $L^1(X_n, \mathcal{S}_n, p_n)$ is weakly Bayes connected if and only if the state space of $L^1(X_m, \mathcal{S}_m, p_m)$ is. Furthermore, if p_n is faithful, then $L^1(X_n, \mathcal{S}_n, p_n) = \mathcal{L}^1(X_n, \mathcal{S}_n, p_n)$ and $\mathbb{E}(\cdot \mid \mathcal{C}) = \mathscr{E}(\cdot \mid \mathcal{C})$ for any $\mathcal{C}$-conditional expectation. Thus one can carry out the calculations involving conditional expectations $\mathbb{E}(\cdot \mid \mathcal{C})$ in terms of the unique version $\mathscr{E}(\cdot \mid \mathcal{C})$. This will be relied on below.

Since X_n is finite, there exist only a finite number of non-trivial Boolean subalgebras $\mathcal{C}_l$ $(l = 1, 2, \ldots, M)$ of $\mathcal{S}_n$; non trivial meaning that $\mathcal{C}_l$ is not $\{\emptyset, X_n\}$ and is not the full Boolean algebra $\mathcal{S}_n$. Each $\mathcal{C}_l$-conditional expectation $\mathscr{E}(\cdot \mid \mathcal{C}_l)$ has the form (cf. Proposition 3.3)

$$\mathscr{E}(\chi_B \mid \mathcal{C}_l) = \sum_k^K \frac{p_n(A_k^l \cap B)}{p_n(A_k^l)} \chi_{A_k^l}, \tag{66}$$

where (for any fixed l) A_k^l $(k = 1, 2, \ldots K)$ is a partition of $\mathcal{S}_n$ and χ_B is the characteristic function of $B \in \mathcal{S}_n$. Assume that $\psi \overset{\mathbb{E}}{\rightsquigarrow} \phi$. Then for some $\mathcal{C}_l$ we have

$$\phi(\chi_B) = \psi(\mathscr{E}(\chi_B \mid \mathcal{C}_l)) \qquad \text{for all } B \in \mathcal{S}_n. \tag{67}$$

Since $\mathcal{C}_l$ is a non-trivial Boolean subalgebra of $\mathcal{S}_n$, at least one A^l_k in $\mathcal{C}_l$ has more than one element from X_n; so if

$$A^l_k = \{x^l_{k_1}, x^l_{k_2}, \dots x^l_{k_l}\} \tag{68}$$

then there exist two distinct elements $x^l_{k_1}, x^l_{k_2}$ in A^l_k. Using (66) and keeping in mind that A^l_k form a partition, we can calculate the probabilities $\phi(\chi_{\{x^l_{k_1}\}})$ and $\phi(\chi_{\{x^l_{k_2}\}})$ as follows:

$$\phi(\chi_{\{x^l_{k_1}\}}) = \psi(\mathscr{E}(\chi_{\{x^l_{k_1}\}} \mid \mathcal{C}_l)) \tag{69}$$

$$= \psi(\sum_k^K \frac{p_n(A^l_k \cap \{x^l_{k_1}\})}{p_n(A^l_k)} \chi_{A^l_k}) \tag{70}$$

$$= \psi(\frac{p_n(\{x^l_{k_1}\})}{p_n(A^l_k)} \chi_{A^l_k}) \tag{71}$$

$$= \frac{p_n(\{x^l_{k_1}\})}{p_n(A^l_k)} \psi(\chi_{A^l_k}). \tag{72}$$

Clearly, $\phi(\chi_{\{x^l_{k_2}\}})$ can be calculated exactly the same way and we obtain

$$\phi(\chi_{\{x^l_{k_2}\}}) = \frac{p_n(\{x^l_{k_2}\})}{p_n(A^l_k)} \psi(\chi_{A^l_k}). \tag{73}$$

Equations (69)–(72) and (73) entail that if $\psi \overset{\mathbb{E}}{\rightsquigarrow} \phi$ with respect to the conditional expectation $\mathscr{E}(\cdot \mid \mathcal{C}_l)$, then there exist elements $x^l_{k_1} \neq x^l_{k_2}$ such that

$$\phi(\chi_{\{x^l_{k_1}\}}) = p_n(\{x^l_{k_1}\}) \frac{\psi(\chi_{A^l_i})}{p_n(A^l_k)} \tag{74}$$

$$\phi(\chi_{\{x^l_{k_2}\}}) = p_n(\{x^l_{k_2}\}) \frac{\psi(\chi_{A^l_i})}{p_n(A^l_k)}. \tag{75}$$

It follows (recall that p_n is faithful) that if ϕ is such that

$$\frac{\phi(\chi_{\{x_i\}})}{p_n(\{x_i\})} \neq \frac{\phi(\chi_{\{x_j\}})}{p_n(\{x_j\})} \qquad i \neq j;\ 1 \leq i, j \leq n \tag{76}$$

then $\psi \overset{\mathbb{E}}{\rightsquigarrow} \phi$ cannot hold for any of the finite number of conditional expectations $\mathcal{C}_l$.

That for any faithful p_n there exists a ϕ for which (76) holds follows from the following

LEMMA 6.9. *Let $a_1, a_2, \dots, a_n$ be real numbers in the semi-closed interval* $(0, 1]$ *such that $\sum_i^n a_i = 1$. Then there exist real numbers $b_1, b_2, \dots, b_n$ such that*

$$b_i \in (0, 1] \qquad i = 1, 2, \dots, n \tag{77}$$

$$\sum_i^n b_i = 1 \tag{78}$$

$$\frac{b_i}{a_i} \neq \frac{b_j}{a_j} \quad \textit{for all } i \neq j;\ i, j = 1, 2, \dots, n \tag{79}$$

Proof of Lemma. Simple induction: The case $n = 2$ is trivial. Assume (induction hypothesis) that Lemma is true for $n > 2$. Let $a_1, a_2, \dots, a_{n+1}$ be numbers in $(0, 1]$ such that $\sum_i^{n+1} a_i = 1$. Consider the numbers a'_i defined by

$$a_i' \doteq \frac{a_i}{\sum_i^n a_i} \qquad i = 1, 2, \ldots, n. \tag{80}$$

Then $a_i' \in (0, 1]$, and $\sum_i^n a_i' = 1$, so by the induction hypothesis there exist numbers $b_i \in (0, 1]$ $(i = 1, 2, \ldots, n)$ such that

$$\sum_i^n b_i = 1 \tag{81}$$

$$\frac{b_i}{a_i'} \neq \frac{b_j}{a_j'} \quad \text{for all } i \neq j;\ i, j = 1, 2, \ldots, n. \tag{82}$$

Which entails

$$\frac{b_i}{a_i' \sum_i^n a_i} \neq \frac{b_j}{a_j' \sum_i^n a_i} \quad \text{for all } i \neq j;\ i, j = 1, 2, \ldots, n \tag{83}$$

Hence

$$\frac{b_i}{a_i} \neq \frac{b_j}{a_j} \quad \text{for all } i \neq j;\ i, j = 1, 2, \ldots, n \tag{84}$$

Let

$$M = \max_i \left\{ \frac{b_i}{a_i} : i = 1, 2, \ldots n \right\} \tag{85}$$

and choose b_{n+1} such that $\frac{b_{n+1}}{a_{n+1}} > M$. Then

$$\frac{b_i}{a_i} \neq \frac{b_j}{a_j} \quad \text{for all } i \neq j;\ i, j = 1, 2, \ldots, n+1. \tag{86}$$

Re-normalizing b_i $(i =, 1, 2, \ldots n+1)$ by dividing each b_i by $\sum_i^{n+1} b_i$ in order to satisfy $\sum_i^{n+1} b_i = 1$ preserves (86). So the claim of Lemma is proved. □

The proof of Proposition 6.8 also reveals that there exist in fact a large number of probability measures over a finite Boolean algebra that are Bayes inaccessible: There is not only one state ϕ in $L^1(X_n, \mathcal{S}_n, p_n)^\sharp$ which satisfies equation (76) and hence is not Bayes accessible from any other state: For all small enough numbers ϵ the states ϕ_ϵ such that

$$|\phi_\epsilon(\chi_{\{x_i\}}) - \phi(\chi_{\{x_i\}})| \leq \epsilon \qquad \text{for all, } i = 1, 2, \ldots n \tag{87}$$

also satisfy (76) and thus cannot be obtained via nontrivial conditionalization using conditional expectations from any other state. Thus, we have:

PROPOSITION 6.10. *Given a fixed probability measure representing the background degree of belief of the Bayesian Agent on a finite Boolean algebra, there exist an uncountably infinite number of states that are not Bayes accessible for the Bayesian Agent.*

A similar proposition can be stated for all *standard* probability spaces, as well: the proof of Propositions 6.4 and 6.7 reveals that the functions $f \in L^\infty(X, \mathcal{S}, p) \cap L^2(X, \mathcal{S}, p)$ witnessing non-weak Bayes connectedness of the state spaces $L^s(X, \mathcal{S}, p)^\sharp$ $(s = 1, 2)$ can be chosen infinitely many different ways. This leads to the next proposition.

PROPOSITION 6.11. *Given a fixed probability measure representing the background degree of belief of the Bayesian Agent on a standard probability space, there exist an uncountably infinite number of states that are not Bayes accessible for the Bayesian Agent.*

To sum up: Lack of weak Bayes connectedness of typical state spaces means that there exist probabilities on σ-algebras that are not Bayes accessible for the Bayesian Agent in the given framework: Given the Agent's background degree of belief on the fixed set of propositions, the Agent cannot infer all probability measures via a Bayesian upgrading (using conditional expectations as conditioning device) no matter what evidence he is provided with—if by evidence is meant specifying a probability measure on some proper nontrivial σ-subalgebra of the fixed set of all propositions. This shows the limits of Bayesian learning under the condition that the evidence available for the Agent is restricted to probability measures on σ-subalgebras of a fixed Boolean σ-algebra. Call this Restricted Evidence Upgrading. Given the limits of Bayesian learning as displayed by Propositions 6.8, 6.7 and 6.4 characterizing Restricted Evidence Upgrading, one can ask if the Bayesian Agent can go beyond these limits if the available evidence is not restricted to probability measures on σ-subalgebras of a fixed σ-algebra. This issue will be investigated in §8. The next section deals with the problem of transitivity of the Bayes accessibility relation.

§7. The Bayes accessibility relation is not transitive.

PROPOSITION 7.1. *The Bayes accessibility relation* $\overset{\mathbb{E}}{\rightsquigarrow}$ *on* $L^1(X, \mathcal{S}, p)^\sharp$ *is not transitive if the* σ*-algebra* $\mathcal{S}$ *has more than 4 elements.*

Proof. We show that there exist three states ψ, ϕ, and ρ in $L^1(X, \mathcal{S}, p)^\sharp$ such that $\psi \overset{\mathbb{E}}{\rightsquigarrow} \phi$ and $\phi \overset{\mathbb{E}}{\rightsquigarrow} \rho$ hold but $\psi \overset{\mathbb{E}}{\not\rightsquigarrow} \rho$. (After this proof, an explicit elementary example of such states will be given, see Example 7.2.)

Let $\mathcal{A}$ and $\mathcal{B}$ be two σ-subalgebras of $\mathcal{S}$ such that there exist elements $A \in \mathcal{A} \setminus \mathcal{B}$ and $B \in \mathcal{B} \setminus \mathcal{A}$ such that $A \cap B = C \neq \emptyset$. Note that if $\mathcal{S}$ has more than four elements, then there exist σ-subalgebras $\mathcal{A}$ and $\mathcal{B}$ of $\mathcal{S}$ with this property: If $\mathcal{S}$ has more than 4 elements, then it has at least 8 elements, and thus there are elements A and B lying in a general position; that is to say, there exist elements A and B for which the following conditions hold:

$$A \not\subseteq B, \quad B \not\subseteq A, \quad A \cap B \neq \emptyset, \quad A \cup B \neq X \tag{88}$$

Let $\mathcal{A}$ and $\mathcal{B}$ be the σ-subalgebras generated by A and B, respectively

$$\mathcal{A} = \{\emptyset, A, A^\perp, X\}, \qquad \mathcal{B} = \{\emptyset, B, B^\perp, X\}. \tag{89}$$

Then $\mathcal{A} \neq \mathcal{B}$, and A and B with the assumed property exist.

Let $\mathbb{E}_\mathcal{A}$ and $\mathbb{E}_\mathcal{B}$ be the two projections on the Hilbert space $L^2(X, \mathcal{S}, p)$ corresponding to the $\mathcal{A}$-conditional and $\mathcal{B}$-conditional expectations $\mathbb{E}(\cdot \mid \mathcal{A})$ and $\mathbb{E}(\cdot \mid \mathcal{B})$, respectively. The set of all projections on $L^2(X, \mathcal{S}, p)$ form an orthocomplemented, orthomodular lattice (see e.g., [28, 38]), where orthomodularity is the property that for any two projections Q and R one has

$$\text{if } Q \leq R \text{ then } R = Q \vee (R \wedge Q^\perp). \tag{90}$$

Applying (90) to $Q = [\mathbb{E}_\mathcal{A} \wedge \mathbb{E}_\mathcal{B}]$ and $R = \mathbb{E}_\mathcal{A}$ and $R = \mathbb{E}_\mathcal{B}$, we obtain

$$\mathbb{E}_\mathcal{A} = [\mathbb{E}_\mathcal{A} \wedge \mathbb{E}_\mathcal{B}] \vee (\mathbb{E}_\mathcal{A} \wedge [\mathbb{E}_\mathcal{A} \wedge \mathbb{E}_\mathcal{B}]^\perp). \tag{91}$$

$$\mathbb{E}_\mathcal{B} = [\mathbb{E}_\mathcal{A} \wedge \mathbb{E}_\mathcal{B}] \vee (\mathbb{E}_\mathcal{B} \wedge [\mathbb{E}_\mathcal{A} \wedge \mathbb{E}_\mathcal{B}]^\perp). \tag{92}$$

Since $[\mathbb{E}_\mathcal{A} \wedge \mathbb{E}_\mathcal{B}]$ is orthogonal to both $(\mathbb{E}_\mathcal{A} \wedge [\mathbb{E}_\mathcal{A} \wedge \mathbb{E}_\mathcal{B}]^\perp)$ and to $(\mathbb{E}_\mathcal{B} \wedge [\mathbb{E}_\mathcal{A} \wedge \mathbb{E}_\mathcal{B}]^\perp)$, and since the join of orthogonal projections is equal to their sum, equations (91)–(92) can be written as

$$\mathbb{E}_{\mathcal{A}} = [\mathbb{E}_{\mathcal{A}} \wedge \mathbb{E}_{\mathcal{B}}] + (\mathbb{E}_{\mathcal{A}} \wedge [\mathbb{E}_{\mathcal{A}} \wedge \mathbb{E}_{\mathcal{B}}]^{\perp}). \tag{93}$$

$$\mathbb{E}_{\mathcal{B}} = [\mathbb{E}_{\mathcal{A}} \wedge \mathbb{E}_{\mathcal{B}}] + (\mathbb{E}_{\mathcal{B}} \wedge [\mathbb{E}_{\mathcal{A}} \wedge \mathbb{E}_{\mathcal{B}}]^{\perp}). \tag{94}$$

The product $\mathbb{E}_{\mathcal{A}}\mathbb{E}_{\mathcal{B}}$ is a projection if and only if $\mathbb{E}_{\mathcal{A}}$ and $\mathbb{E}_{\mathcal{B}}$ commute as operators. Relations (93)–(94) show that $\mathbb{E}_{\mathcal{A}}$ and $\mathbb{E}_{\mathcal{B}}$ commute if and only if their parts outside their intersection are orthogonal, i.e., if and only if $(\mathbb{E}_{\mathcal{A}} \wedge [\mathbb{E}_{\mathcal{A}} \wedge \mathbb{E}_{\mathcal{B}}]^{\perp})$ and $(\mathbb{E}_{\mathcal{B}} \wedge [\mathbb{E}_{\mathcal{A}} \wedge \mathbb{E}_{\mathcal{B}}]^{\perp})$ are orthogonal. But $(\mathbb{E}_{\mathcal{A}} \wedge [\mathbb{E}_{\mathcal{A}} \wedge \mathbb{E}_{\mathcal{B}}]^{\perp})$ and $(\mathbb{E}_{\mathcal{B}} \wedge [\mathbb{E}_{\mathcal{A}} \wedge \mathbb{E}_{\mathcal{B}}]^{\perp})$ are *not* orthogonal because by assumption there exist elements $A \in \mathcal{A}$ and $B \in \mathcal{B}$ such that $A \notin \mathcal{B}$ and $B \notin \mathcal{A}$, so the characteristic functions χ_A and χ_B of the elements $A \in \mathcal{A}$ and $B \in \mathcal{B}$ are in the range of the projections $(\mathbb{E}_{\mathcal{A}} \wedge [\mathbb{E}_{\mathcal{A}} \wedge \mathbb{E}_{\mathcal{B}}]^{\perp})$ and $(\mathbb{E}_{\mathcal{B}} \wedge [\mathbb{E}_{\mathcal{A}} \wedge \mathbb{E}_{\mathcal{B}}]^{\perp})$, respectively, and by the condition $A \cap B = C \neq 0$, for the L^2 scalar product $\langle \chi_A, \chi_B \rangle$ of χ_A and χ_B we have

$$\langle \chi_A, \chi_B \rangle = \int_X \chi_A \chi_B dp = \int_X \chi_C dp = p(C) \neq 0, \tag{95}$$

where we used that p is faithful on $L^2(X, \mathcal{S}, p)$.

Since the product $\mathbb{E}_{\mathcal{A}}\mathbb{E}_{\mathcal{B}}$ is not a projection, it is not equal to any projection $\mathbb{E}_{\mathcal{C}}$ that would represent a conditional expectation $\mathbb{E}(\cdot \mid \mathcal{C})$ defined by a σ-subalgebra $\mathcal{C}$ of $\mathcal{S}$. Thus for any such projection $\mathbb{E}_{\mathcal{C}}$ there is an element $f \in L^2(X, \mathcal{S}, p) \subset L^1(X, \mathcal{S}, p)$ such that

$$\mathbb{E}_{\mathcal{A}}\mathbb{E}_{\mathcal{B}} f \neq \mathbb{E}_{\mathcal{C}} f. \tag{96}$$

The state space $L^1(X, \mathcal{S}, p)^{\sharp}$ is separating: for any $f \neq g$ in $L^1(X, \mathcal{S}, p)$, there exists a state ψ in $L^1(X, \mathcal{S}, p)^{\sharp}$ such that $\psi(f) \neq \psi(g)$, so there is a state ψ such that

$$\psi(\mathbb{E}_{\mathcal{A}}\mathbb{E}_{\mathcal{B}} f) \neq \psi(\mathbb{E}_{\mathcal{C}} f). \tag{97}$$

It follows that defining states ϕ and ρ by

$$\phi(f) \doteq \psi(\mathbb{E}(f \mid \mathcal{A})) \tag{98}$$

$$\rho(f) \doteq \phi(\mathbb{E}(f \mid \mathcal{B})) \tag{99}$$

we have $\psi \overset{\mathbb{E}}{\rightsquigarrow} \phi$ and $\phi \overset{\mathbb{E}}{\rightsquigarrow} \rho$ but $\psi \overset{\mathbb{E}}{\rightsquigarrow} \rho$ does not hold. □

We illustrate failure of transitivity of the Bayes accessibility relation with the following example.

EXAMPLE 7.2. *We give an explicit example of a probability space and three states ϕ, ψ and ρ in its state space such that $\phi \overset{\mathbb{E}}{\rightsquigarrow} \psi$, $\psi \overset{\mathbb{E}}{\rightsquigarrow} \rho$ but $\phi \overset{\mathbb{E}}{\rightsquigarrow} \rho$ does not hold.*

Let $X_3 = \{x_1, x_2, x_3\}$, $\mathcal{S}_3$ be the power set of X_3, and p_3 be the uniform measure on X_3: $p_3(\{x_i\}) = \frac{1}{3}$ $(i = 1, 2, 3)$. There are three nontrivial Boolean subalgebras of $\mathcal{S}$, they are:

$$\mathcal{C}_1 = \{\emptyset, \{x_1\}, \{x_2, x_3\}, X_3\} \tag{100}$$

$$\mathcal{C}_2 = \{\emptyset, \{x_2\}, \{x_1, x_3\}, X_3\} \tag{101}$$

$$\mathcal{C}_3 = \{\emptyset, \{x_3\}, \{x_1, x_2\}, X_3\} \tag{102}$$

$\mathbb{E}(\cdot \mid \mathcal{C}_1)$, $\mathbb{E}(\cdot \mid \mathcal{C}_2)$ *and* $\mathbb{E}(\cdot \mid \mathcal{C}_3)$ *are the three conditional expectations from* $L^1(X_3, \mathcal{S}_3, p_3)$ *to* $L^1(X_3, \mathcal{C}_i, p_3)$ $(i = 1, 2, 3)$. *These conditional expectations are given on the characteristic functions χ_B of $B \in \mathcal{S}$ by*

$$\mathbb{E}(\chi_B \mid \mathcal{C}_1) = \frac{p(\{x_1\} \cap B)}{p(\{x_1\})}\chi_{\{x_1\}} + \frac{p(\{x_2, x_3\} \cap B)}{p(\{x_2, x_3\})}\chi_{\{x_2,x_3\}}. \tag{103}$$

$$\mathbb{E}(\chi_B \mid \mathcal{C}_2) = \frac{p(\{x_2\} \cap B)}{p(\{x_2\})}\chi_{\{x_2\}} + \frac{p(\{x_1, x_3\} \cap B)}{p(\{x_1, x_3\})}\chi_{\{x_1,x_3\}}. \tag{104}$$

$$\mathbb{E}(\chi_B \mid \mathcal{C}_3) = \frac{p(\{x_3\} \cap B)}{p(\{x_3\})}\chi_{\{x_3\}} + \frac{p(\{x_1, x_2\} \cap B)}{p(\{x_1, x_2\})}\chi_{\{x_1,x_2\}}. \tag{105}$$

Let ϕ be the state on $L^1(X_3, \mathcal{S}_3, p_3)$ defined by the following probabilities:

$$\phi(\chi_{\{x_1\}}) \doteq \frac{1}{2} \qquad \phi(\chi_{\{x_2\}}) \doteq \frac{1}{6} \qquad \phi(\chi_{\{x_3\}}) \doteq \frac{2}{6}. \tag{106}$$

Let ψ and ρ be the states on $L^1(X_3, \mathcal{S}_3, p_3)$ which are defined by

$$\psi(f) \doteq \phi(\mathbb{E}(f \mid \mathcal{C}_1) \tag{107}$$

$$\rho(f) \doteq \psi(\mathbb{E}(f \mid \mathcal{C}_2). \tag{108}$$

So $\phi \overset{\mathbb{E}}{\rightsquigarrow} \psi$ and $\psi \overset{\mathbb{E}}{\rightsquigarrow} \rho$ hold by the very definition of these states. We claim that $\phi \overset{\mathbb{E}}{\rightsquigarrow} \rho$ does not hold however. To see this, one can explicitly compute the values of ρ, they are:

$$\rho(\chi_{\{x_1\}}) = \frac{3}{8} \qquad \rho(\chi_{\{x_2\}}) = \frac{1}{4} \qquad \rho(\chi_{\{x_3\}}) = \frac{3}{8} \tag{109}$$

One also can compute explicitly the values of $\phi(\mathbb{E}(\chi_B \mid \mathcal{C}_i))$, for $B \in L^1(X_3, \mathcal{S}_3, p_3)$ for each $\mathcal{C}_i$-conditional expectation $i = 1, 2, 3$: For the elementary event $B = \{x_1\}$ these values are:

$$\phi(\mathbb{E}(\chi_{\{x_1\}} \mid \mathcal{C}_1)) = \frac{1}{2} \tag{110}$$

$$\phi(\mathbb{E}(\chi_{\{x_1\}} \mid \mathcal{C}_2)) = \frac{5}{12} \tag{111}$$

$$\phi(\mathbb{E}(\chi_{\{x_1\}} \mid \mathcal{C}_3)) = \frac{1}{3} \tag{112}$$

Thus

$$\phi(\mathbb{E}(\chi_{\{x_1\}} \mid \mathcal{C}_i)) \neq \rho(\chi_{\{x_1\}}) \qquad \textit{for all } i = 1, 2, 3 \tag{113}$$

and since $\mathcal{C}_i$, $(i = 1, 2, 3)$ are the only nontrivial Boolean subalgebras of $\mathcal{S}_3$, on can conclude that $\phi \overset{\mathbb{E}}{\rightsquigarrow} \rho$ does not hold.

Lack of transitivity of the Bayes accessibility relation makes the following definition of *finite* Bayes accessibility nonredundant:

DEFINITION 7.3. *A state $\phi \in L^1(X, \mathcal{S}, p)^\sharp$ is called* finitely *Bayes accessible from state ψ if there is a natural number N and there exist σ-subalgebras $\mathcal{C}_i$ of $\mathcal{S}$ $(i = 1, 2, \ldots N)$ such that for the corresponding conditional expectations $\mathbb{E}(\cdot \mid \mathcal{C}_i)$ $(i = 1, 2, \ldots N)$ we have*

$$\phi = \psi \circ \mathbb{E}(\cdot \mid \mathcal{C}_N) \circ \mathbb{E}(\cdot \mid \mathcal{C}_{N-1}) \circ \cdots \circ \mathbb{E}(\cdot \mid \mathcal{C}_1). \tag{114}$$

The interpretation of finite Bayes accessibility of ϕ is that an Agent can genuinely Bayes-learn from error and feedback: the Agent can learn the finitely Bayes accessible state ϕ from some evidence ψ in N steps. In step i the state $\phi_i \doteq \psi \circ \mathbb{E}(\cdot \mid \mathcal{C}_{i-1}) \circ \cdots \circ \mathbb{E}(\cdot \mid \mathcal{C}_1)$ inferred in the preceding $(i-1)th$ step is looked at and it is confirmed that state ϕ_i is correct on the subspace $L^1(X, \mathcal{C}_i, p_{\mathcal{C}_i})^\sharp$; hence the restriction of ϕ_i to this subspace should be considered as new evidence and a new state ϕ_{i+1} should be inferred on this basis by conditionalizing (extending) ϕ_i from $L^1(X, \mathcal{C}_i, p_{\mathcal{C}_i})^\sharp$ to $L^1(X, \mathcal{S}, p)^\sharp$ using the conditional expectation $\mathbb{E}(\cdot \mid \mathcal{C}_i)$. At the N-th step ϕ is learned. All this learning is taking place while the probability measure p representing the Agent's background knowledge is kept fixed. From this perspective of finite Bayes learning, failure of transitivity of $\overset{\mathbb{E}}{\rightsquigarrow}$ means that "There is no Bayesian royal road to learning" in general: Even if a state can be learned from from an initial evidence by performing several successive steps of conditionalizing based on feedback, this step-by-step learning cannot be shortcut in general by a single Bayesian learning move.

Finite Bayes accessibility is relevant in situations where the Agent is receiving (partial and revisable) probabilistic information piece by piece at certain times over a period of time. As a hypothetical example consider a situation when the distribution of occurrence of specific attributes in a large data set should be learned by a Bayesian learning machine characterized by its prior probability distribution (background). The data analysis producing evidence for the Bayes-learning machine can be a process such that only the joint occurrence of certain attributes can be extracted from the data (i.e., only the distribution of elements in a nontrivial partition are empirically accessible) and the data analysis can be taking place simultaneously with the machine Bayes-learning from the variable data (evidence) supplied at successive times. The learning dynamic in this situation is described by the learning steps defining finite Bayes accessibility.

The notion of finite Bayes accessibility leads naturally to the question of whether state spaces are *finitely* strongly or weakly Bayes connected in the sense of the following definition, which is the complete analogue of Definition 6.1:

DEFINITION 7.4. *The state space $L^1(X, \mathcal{S}, p)^\sharp$ is said to be*

- finitely strongly Bayes connected *if any state $\phi \in L^1(X, \mathcal{S}, p)^\sharp$ is finitely Bayes accessible from any state $\psi \in L^1(X, \mathcal{S}, p)^\sharp$;*
- finitely weakly Bayes connected *if for any state $\phi \in L^1(X, \mathcal{S}, p)^\sharp$ there exists state $\psi \in L^1(X, \mathcal{S}, p)^\sharp$ such that ϕ is finitely Bayes accessible from ψ, and $\phi \neq \psi$.*

Finite strong/weak Bayes connectedness is a weakening of strong/weak Bayes connectedness; thus, in principle, it is easier for a state space to be finitely strongly/weakly Bayes connected than just being strongly/weakly Bayes connected. It is clear however that if state ϕ is Bayes inaccessible then it also is finitely Bayes inaccessible because if it is finitely Bayes accessible then it can be obtained in the form (114) hence it is Bayes accessible from $\psi \circ \mathbb{E}(\cdot \mid \mathcal{C}_{N-1}) \circ \cdots \circ \mathbb{E}(\cdot \mid \mathcal{C}_1)$ and thus it is Bayes accessible. So we have

PROPOSITION 7.5. *If a state space is not weakly Bayes connected then it is not* finitely *weakly Bayes connected.*

As a corollary of Propositions 6.7 and 7.4 we have in particular:

PROPOSITION 7.6. *State spaces of standard probability measure spaces are not finitely weakly Bayes connected (hence they are not finitely strongly Bayes connected either).*

Though standard probability spaces include all finite spaces, to explore the behavior of finite weak Bayes connectedness it is illustrative to provide a direct proof of failure of finite strong Bayes connectedness in probability spaces having a finite number of events. This is our next proposition.

PROPOSITION 7.7. *Suppose $\mathcal{S}$ is finite. Then $L^1(X, \mathcal{S}, p)^\sharp$ is not finitely strongly Bayes connected.*

Proof. Suppose $\mathcal{C}$ is a σ-subalgebra. Then it is generated by a partition $(A_k)_{k\in n_{\mathcal{C}}} \subseteq X$ and by Proposition 3.3 the $\mathcal{C}$-conditional expectation $\mathbb{E}(\chi_B \mid \mathcal{C})$ is given by

$$\mathbb{E}(\chi_B \mid \mathcal{C}) = \sum_i \frac{p(B\cap A_i)}{p(A_i)}\chi_{A_i} \quad (B\in\mathcal{S}) \tag{115}$$

(where in case of $p(A_i) = 0$ the coefficient of χ_{A_i} can be set arbitrarily. As conditional expectations are unique up to measure zero, we may assume without loss of generality that in case $p(A_i) = 0$ the coefficient of χ_{A_i} is 0). Each measurable $f : X \to \mathbb{R}$ is a sum of characteristic functions

$$f = \sum_{x\in X} f(x)\chi_{\{x\}}$$

therefore if a state is defined on all $\chi_{\{x\}}$, then it extends uniquely to $L^1(X, \mathcal{S}, p)$.

Enumerate the range of p. As $\mathcal{S}$ is finite, our enumeration is finite as well and we can write the range $\mathrm{ran}(p)$ of p as

$$\mathrm{ran}(p) = \left\{r_k : \ 1 \le k \le K\right\}.$$

Consider the field extension $\mathbb{F} = \mathbb{Q}[r_1, \ldots, r_K]$ (for the theory of field extensions we refer to [40]). It is clear that all the coefficients in equation (115) belong to the field $\mathbb{F}$. Since $\mathrm{ran}(p)$ is finite, we have

$$\mathbb{Q}[r_1, \ldots, r_K] \neq \mathbb{R}.$$

Observe that if ψ takes values from $\mathbb{F}$ on all $\chi_{\{x\}}$, then the same holds for $\psi\circ\mathbb{E}(\cdot \mid \mathcal{C})$ for any $\mathcal{C}$, since all the coefficients in equation (115) are in $\mathbb{F}$.

Choose now any ψ which takes values from $\mathbb{F}$ on all the $\chi_{\{x\}}$'s ($x \in X$) and suppose ϕ takes values from $\mathbb{R}\setminus\mathbb{F}$ on each $\chi_{\{x\}}$. Then

$$\phi(\chi_{\{x\}}) \neq \psi\circ\mathbb{E}(\chi_{\{x\}} \mid \mathcal{C}_N)\circ\mathbb{E}(\chi_{\{x\}} \mid \mathcal{C}_{N-1})\circ\cdots\circ\mathbb{E}(\chi_{\{x\}} \mid \mathcal{C}_1),$$

for any N, since the right hand side of the equation belongs to $\mathbb{F}$, while the left hand side is in $\mathbb{R}\setminus\mathbb{F}$. □

As the proof of Proposition 7.1 shows, the failure of transitivity of the Bayes accessibility relation is a consequence of noncommutativity of the projections $\mathbb{E}_{\mathcal{A}}$ and $\mathbb{E}_{\mathcal{B}}$ representing the conditional expectations determined by *some* σ-subalgebras $\mathcal{A}$ and $\mathcal{B}$ of σ-algebra $\mathcal{S}$. It is not true however that all subalgebras $\mathcal{A}$ and $\mathcal{B}$ of $\mathcal{S}$ determine noncommuting projections $\mathbb{E}_{\mathcal{A}}$ and $\mathbb{E}_{\mathcal{B}}$. For instance, if $\mathcal{A}$ is a sub-Boolean algebra of $\mathcal{B}$, then $\mathbb{E}_{\mathcal{A}}$ is a sub-projection of $\mathbb{E}_{\mathcal{B}}$ and these projections commute. More generally, if $\mathcal{C}_i$ ($i = 1, 2, \ldots$) is a series of σ-subalgebras of $\mathcal{S}$ such that $\mathcal{C}_i \subset \mathcal{C}_j$ for $i < j$ (such a series is called a "filtration" [1, p. 458]) then by the "tower property" of conditional expectations [1, Theorem 34.4] we have

$$\mathbb{E}(\mathbb{E}(\cdot \mid \mathcal{C}_j) \mid \mathcal{C}_i) = \mathbb{E}(\cdot \mid \mathcal{C}_i) \quad (i < j) \tag{116}$$

which in turn entails the commutativity of the Hilbert space projections representing $\mathbb{E}(\cdot \mid \mathcal{C}_j)$ and $\mathbb{E}(\cdot \mid \mathcal{C}_i)$. Thus in any such particular series of Bayesian learning steps transitivity holds. Furthermore, if $\mathcal{B}$ is a subalgebra such that $\mathcal{B} = \cup_i\mathcal{C}_i$, then the (upward) martingale theorem (Theorem 35.6 in [1]) says that for any ψ in $L^1(X, \mathcal{S}, p)^\sharp$ we have

$$\psi\circ\mathbb{E}(\cdot \mid \mathcal{B}) = \lim_i \psi\circ(\mathbb{E}(\cdot \mid \mathcal{C}_i)). \tag{117}$$

(The limit is to be taken in the pointwise topology in the state space: $\phi_n \to \phi$ if $\phi_n(f) \to \phi(f)$ for all $f \in L^1(X, \mathcal{S}, p)$.) Performing repeated conditionalization with respect to subalgebras forming a filtration describes a Bayesian learning process in which the inferred probabilities in the first step using conditionalization with respect to Boolean algebra $\mathcal{C}_1$ are confirmed as correct (for instance as describing empirically observed frequencies) on all larger algebras $\mathcal{C}_i$ $(i > 1)$, because the tower property (116) entails that the finite Bayes accessibility relation defined by equation (114) reduces to

$$\phi = \psi \circ \mathbb{E}(\cdot \mid \mathcal{C}_N) \circ \mathbb{E}(\cdot \mid \mathcal{C}_{N-1}) \circ \cdots \circ \mathbb{E}(\cdot \mid \mathcal{C}_1) = \psi \circ \mathbb{E}(\cdot \mid \mathcal{C}_1). \tag{118}$$

The martingale equation (117) also can be interpreted as a formulation of the phenomenon known as "washing out of priors" (Chapter 6, sec. 4 in [8]), and this can be used to clarify the relation of this phenomenon to the Bayes inaccessibility of a state for an Agent: If the filtration covers the whole $\mathcal{S}$, i.e., $\mathcal{S} = \cup_i \mathcal{C}_i$, then, as a special instance of equation (117) we have:

$$\psi = \lim_i \psi \circ (\mathbb{E}(\cdot \mid \mathcal{C}_i)). \tag{119}$$

In equation (119) all the conditional expectations $\mathbb{E}(\cdot \mid \mathcal{C}_i)$ are defined with respect to the background probability (prior) p that determines the function space $L^1(X, \mathcal{S}, p)$ on which $\mathbb{E}(\cdot \mid \mathcal{C}_i)$ act. One can however take another probability measure q and form $L^1(X, \mathcal{S}, q)$. Then, *if* state ψ is such that it also belongs to the state space $L^1(X, \mathcal{S}, q)^\sharp$, then the filtration $\mathcal{C}_i$ defines the series of conditional expectations $\mathbb{E}_q(\cdot \mid \mathcal{C}_i)$ on $L^1(X, \mathcal{S}, q)$, and the martingale theorem applied to this series of conditional expectations yields

$$\psi = \lim_i \psi \circ (\mathbb{E}_q(\cdot \mid \mathcal{C}_i)). \tag{120}$$

Thus either p or q can be taken as prior to obtain ψ as a limit of the conditioned probabilities $\psi \circ (\mathbb{E}_q(\cdot \mid \mathcal{A}_i))$ provided $\mathcal{C}_i$ is a filtration covering (in the limit) the whole algebra $\mathcal{S}$—the prior "gets washed out" in the limit. This holds for any ψ that belongs to both $L^1(X, \mathcal{S}, p)^\sharp$ and $L^1(X, \mathcal{S}, q)^\sharp$; in particular this holds for a ψ that is *not* Bayes accessible for an Agent having p as his background measure and which is *not* Bayes accessible for an Agent having q as background either. Recall (Definition 4.1) that the Bayes inaccessibility for an Agent of ψ in $L^1(X, \mathcal{S}, p)^\sharp$ and $L^1(X, \mathcal{S}, q)^\sharp$ means that

$$\psi \neq \psi \circ (\mathbb{E}(\cdot \mid \mathcal{C}_i)) \quad \text{for any } i \text{ such that } \mathcal{C}_i \subset \mathcal{S}. \tag{121}$$

$$\psi \neq \psi \circ (\mathbb{E}_q(\cdot \mid \mathcal{A}_i)) \quad \text{for any } i \text{ such that } \mathcal{C}_i \subset \mathcal{S}. \tag{122}$$

That is to say, while ψ can be obtained as a *limit* of both $\psi \circ (\mathbb{E}(\cdot \mid \mathcal{C}_i))$ and $\psi \circ (\mathbb{E}_q(\cdot \mid \mathcal{C}_i))$ if in the limit $\mathcal{C}_i$ covers $\mathcal{S}$, at no stage along the limit can we obtain ψ if at that stage the conditionalization is with respect to a *proper* subalgebra of $\mathcal{S}$.

It should be noted that the right hand sides of the martingale equations (119) and (120) represent a series of individual, disconnected Bayesian inferences rather than a connected Bayesian learning process as understood in the concept of finite Bayes accessibility: every conditionalization $\psi \circ \mathbb{E}(\cdot \mid \mathcal{C}_i)$ and $\psi \circ \mathbb{E}_q(\cdot \mid \mathcal{C}_i)$ is a Bayesian inference that infers probabilities of events in $\mathcal{S}$ from probabilities on $\mathcal{C}_i$ (with respect to priors p and q, respectively). But the result of the i-th inference does not play any role in the $(i + 1)-$th inference in this series, in contrast to the series of inferences in the concept of finite Bayes accessibility.

One also should be aware of the limitation of the washing out of prior phenomenon resulting from the hypothesis under which it holds: Given a state ψ to learn, the probability measures p and q cannot be arbitrary: They must be such that ψ belongs to *both* of the state spaces $L^1(X, \mathcal{S}, p)^\sharp$ and $L^1(X, \mathcal{S}, q)^\sharp$ otherwise equations (119) and (120) are not meaningful. In other words, p and q must be such that the probability measure ψ defines on $\mathcal{S}$ is absolutely continuous with respect to both p and q. This is a very heavy constraint limiting the scope of truth of the claim that priors do not matter in the long run. For further discussion of the limits of washing out of priors we refer to (Chapter 6, sec. 4 in [8]).

§8. Bayes connectability in terms of conditional expectations. Failure of weak Bayes connectedness of state spaces displays the limits of Bayesian learning under Limited Evidence Upgrading: the evidence available for the Agent is limited to probability measures on σ-subalgebras of the fixed σ-algebra on which the Agent's background probability is given. It is natural however to ask what the Agent can learn via conditionalization using conditional expectations if he is allowed access to potentially unlimited evidence. To investigate this question, we define first the concept of extensions of state spaces.

DEFINITION 8.1. *We say that the state space $L^1(X', \mathcal{S}', p')^\sharp$ extends the state space $L^1(X, \mathcal{S}, p)^\sharp$ if the following hold*:

(i) *There is a measurable, measure preserving map $h : X' \to X$ such that its inverse image function f^{-1} induces a σ-algebra embedding $h^{-1} : \mathcal{S} \to \mathcal{S}'$.*

(ii) *The σ-algebra embedding preserves the probability*: *For all $A \in \mathcal{S}$ we have $p(A) = p'(h^{-1}(A))$.*

If (i)–(ii) hold, then the embedding of $\mathcal{S}$ into $\mathcal{S}'$ via f^{-1} can be lifted to an isometric embedding $\bar{h} : L^1(X, \mathcal{S}, p) \to L^1(X', \mathcal{S}', p')$ by defining $\bar{h}$ in the natural way: For a function $f \in L^1(X, \mathcal{S}, p)$ let $\bar{h}(f) = f \circ h$ (see the figure below). Since h is measurable, we have $\bar{h}(f) \in L^1(X', \mathcal{S}', p')$.

$$\begin{tikzcd} X \arrow[r, "f"] & \mathbb{R} \arrow[d, leftrightarrow, "id"] \\ X' \arrow[u, "h"] \arrow[r, dashed, "f \circ h"'] & \mathbb{R} \end{tikzcd}$$

Note that $\bar{h}$ is isometric because

$$\|\bar{h}(f)\|_1 = \int_{X'} |f \circ h| dp' = \int_X |f| dp = \|f\|_1 \tag{123}$$

The image of $\bar{h}$ is thus a closed subspace in $L^1(X', \mathcal{S}', p')$; hence for each state $\phi \in L^1(X, \mathcal{S}, p)^\sharp$ there is a corresponding state $\bar{\phi} \in \bar{h}\big(L^1(X, \mathcal{S}, p)\big)^\sharp$ such that

$$\bar{\phi}\big(\bar{h}(f)\big) = \phi(f) \quad \textit{for all } f \in L^1(X, \mathcal{S}, p) \tag{124}$$

By the Hahn–Banach theorem $\bar{\phi}$ extends to a continuous linear functional $\phi' \in L^1(X', \mathcal{S}', p')^\sharp$. Notice that $\bar{\phi}$ can have many such extensions, in general. Any such ϕ' is called an extension of ϕ.

DEFINITION 8.2. *The state space $L^1(X, \mathcal{S}, p)^\sharp$ is called weakly Bayes* connectable *if there is a state space extension $L^1(X', \mathcal{S}', p')^\sharp$ such that each $\phi \in L^1(X, \mathcal{S}, p)^\sharp$ has an extension $\phi' \in L^1(X', \mathcal{S}', p')^\sharp$ which is Bayes accessible from* some *$\psi \in L^1(X', \mathcal{S}', p')^\sharp$, $\psi \neq \phi'$.*

The above definition is a significantly generalized version of the definition given by Diaconis and Zabell [4, sec. 2.1]. Accordingly, the proposition below generalizes Theorem 2.1 in [4].

PROPOSITION 8.3. *State spaces $L^1(X, \mathcal{S}, p)^\sharp$ are weakly Bayes connectable.*

Proof. Let ϕ be a state in $L^1(X, \mathcal{S}, p)^\sharp$. We have to construct a state space extension $L^1(X', \mathcal{S}', p')^\sharp$ such that the extension $\phi' \in L^1(X', \mathcal{S}', p')^\sharp$ of $\phi \in L^1(X, \mathcal{S}, p)^\sharp$ is Bayes accessible from some $\psi \in L^1(X', \mathcal{S}', p')^\sharp$. The idea of the proof is the following. We take as the extension of $L^1(X, \mathcal{S}, p)$ the product of $L^1(X, \mathcal{S}, p)$ with another probability space $(Y, \mathcal{B}, q)$. The product structure defines a conditional expectation to the components in the product in a canonical manner, and it also makes possible to extend states defined on the components in different ways. We display two extensions of ϕ that will be shown to be related to each other via conditioning with respect to the canonical conditional expectation.

Let $(Y, \mathcal{B}, q)$ be the Lebesgue measure space over the unit interval and consider the usual product space

$$(X \times Y, \mathcal{S} \otimes \mathcal{B}, p \times q), \tag{125}$$

where $p \times q$ is the product measure: $(p \times q)(A \times B) = p(A)q(B)$.

The function

$$X \times Y \ni (x, y) \mapsto h(x, y) \doteq x \in X \tag{126}$$

is a measurable, measure preserving map and its inverse image induces a σ-algebra embedding $h^{-1} : \mathcal{S} \to \mathcal{S}'$, since for all $A \in \mathcal{S}$ we have

$$h^{-1}(A) = A \times Y \in \mathcal{S}' \tag{127}$$
$$p(A) = p'(A \times Y) = p(A)q(Y) \tag{128}$$

h can be lifted to an isometric embedding $\bar{h} : L^1(X, \mathcal{S}, p) \to L^1(X', \mathcal{S}', p')$ by the definition

$$\bar{h}(f) = \bar{f} = f \circ h \qquad f \in L^1(X, \mathcal{S}, p). \tag{129}$$

In what follows, for notational convenience we write $L^1(X)$ and $L^1(X \times Y)$ instead of the longer $L^1(X, \mathcal{S}, p)$ and $L^1(X', \mathcal{S}', p')$, and, to make notation easier to read, we write $\int dx$ and $\int dy$ instead of $\int dp$ and $\int dq$.

The general definition of extension of state spaces (Definition 8.1) in the present context means that if $\phi \in L^1(X)^\sharp$ is a state, then $\phi' \in L^1(X \times Y)^\sharp$ is its extension if for all $f \in L^1(X)$ we have

$$\phi'(\bar{f}) = \phi(f). \tag{130}$$

If $\alpha \in L^1(Y)^\sharp$ is a state in the second component of the product space (125), then we define the α-extension of ϕ (denoted by ϕ_α) to be a state in $L^1(X \times Y)^\sharp$ by setting for any $f \in L^1(X \times Y)$

$$\phi_\alpha(f) \doteq \alpha\big(y \mapsto \phi(x \mapsto f(x, y))\big). \tag{131}$$

Then ϕ_α is an extension of ϕ because for each $f \in L^1(X)$ we have

$$\phi_\alpha(\bar{f}) = \phi_\alpha(f \circ h) = \alpha\big(y \mapsto \phi(x \mapsto (f \circ h)(x, y))\big) \tag{132}$$
$$= \alpha(y \mapsto \phi(x \mapsto f(x))) = \alpha(y \mapsto \phi(f)) \tag{133}$$
$$= \phi(f) \cdot \alpha(y \mapsto 1) = \phi(f) \cdot \alpha(\mathbf{1}) = \phi(f). \tag{134}$$

A particular state α is given by $\alpha(g) = \int_Y g \; dy$ (for all $g \in L^1(Y)$). For this α the α-extension ϕ_α of ϕ is

$$\bar{\phi}(f) = \int_Y \Big(y \mapsto \phi\big(x \mapsto f(x, y)\big)\Big)\, dy. \tag{135}$$

Take the σ-subalgebra $\mathcal{A} = \{A \times Y : A \in \mathcal{S}\}$ of $\mathcal{S} \times \mathcal{B}$ (which is isomorphic to $\mathcal{S}$). Then the $\mathcal{A}$-conditional expectation is

$$\mathbb{E}(f \mid \mathcal{A})(x, y) = \int_Y f(x, y)\, dy. \tag{136}$$

We claim that for any $\alpha \in L^1(Y)^\sharp$ the state $\bar{\phi}$ is Bayes accessible from ϕ_α using the $\mathcal{A}$-conditional expectation as upgrading device; i.e., that we have

$$\bar{\phi}(f) = \phi_\alpha(\mathbb{E}(f \mid \mathcal{A})). \tag{137}$$

To show this, note first that, since the dual space $L^1(X, \mathcal{S}, p)^*$ is $L^\infty(X, \mathcal{S}, p)$ ([27, Theorem 1.7.8]), there is a function $g \in L^\infty(X)$ such that

$$\phi(f) = \int_X f(x)g(x)\, dx \qquad \text{for all } f \in L^1(X). \tag{138}$$

Then for all $f \in L^1(X \times Y)$ we have

$$\bar{\phi}(f) = \int_Y \Big(y \mapsto \phi\big(x \mapsto f(x, y)\big)\Big)\, dy \tag{139}$$

$$= \int_Y \int_X f(x, y)g(x)\, dxdy. \tag{140}$$

Using the formula (136) giving the conditional expectation $\mathbb{E}(\cdot \mid \mathcal{A})$ and changing the order of integrals below (allowed by Fubini's theorem) we can calculate then

$$\phi\big(\mathbb{E}(f \mid \mathcal{A})\big) = \phi\big(x \mapsto \int_Y f(x, y)\, dy\big) = \int_X \int_Y f(x, y)\, dy\; g(x)\, dx \tag{141}$$

$$= \int_X \int_Y f(x, y)g(x)\, dy\, dx = \int_Y \int_X f(x, y)g(x)\, dx\, dy \tag{142}$$

$$= \int_Y \Big(y \mapsto \big(\int_X f(x, y)g(x)\, dx\big)\Big)\, dy \tag{143}$$

$$= \int_Y \Big(y \mapsto \big(x \mapsto \phi(x \mapsto f(x, y)\big)\Big)\, dy \tag{144}$$

$$= \bar{\phi}\big(f(x, y)\big). \tag{145}$$

Using (141) and (145) the claim (i.e., equation (137)) follows easily:

$$\phi_\alpha\big(\mathbb{E}(f \mid \mathcal{A})\big) = \alpha\big(y \mapsto \phi(\mathbb{E}(f \mid \mathcal{A}))\big) \tag{146}$$

$$= \alpha\big(\bar{\phi}(f)\big) \tag{147}$$

$$= \bar{\phi}(f)\alpha(x \mapsto 1) \tag{148}$$

$$= \bar{\phi}(f)\alpha(\mathbf{1}) = \bar{\phi}(f). \tag{149}$$

To complete the proof one has to show that there exists an α in $L^1(Y)^\sharp$ such that $\phi_\alpha \neq \bar{\phi}$. But this is clear: take any continuous, nonconstant function $t : Y \to \mathbb{R}$ for which $\int_Y t(y)\, dy = 1$ and put $\alpha(g) = \int g(y)t(y)\, dy$.

Thus we proved that ϕ has different extensions $\bar{\phi}$ and ϕ_α such that

$$\bar{\phi}(f) = \phi_\alpha\big(\mathbb{E}(f \mid \mathcal{A})\big). \tag{150}$$

□

Proposition 8.3 shows that a Bayesian Agent can learn in principle everything that can be formulated in terms of a probability measure on a fixed σ-algebra—provided the Agent has access to a potentially unlimited supply of evidence. In this sense a Bayesian Agent has unlimited learning capacity.

Note that it is *not* part of our claim that the additional evidence the Agent needs to have in order to learn a Bayes inaccessible state must be formulated in terms of the product extension of the original probability space the proof of Proposition 8.3 uses. Other extensions might very well transform a Bayes inaccessible state into a Bayes learnable one. It is even to be expected that Bayes learnability of a state via extending might depend sensitively on how the Agent extends the original probability space to accommodate new knowledge.

One may wonder whether state spaces are Bayes connectable in a stronger sense than specified by Definition 8.2; i.e., whether it holds that given any pair of states ϕ and ψ in $L^1(X, \mathcal{S}, p)^\sharp$ such that ϕ is not Bayes accessible from ψ there exists an extension in which ϕ is Bayes accessible from ψ. To give the precise definition of strong Bayes connectability of state spaces, we define first the concept of *in principle Bayes accessibility*:

DEFINITION 8.4. *Given a state space $L^1(X, \mathcal{S}, p)^\sharp$, a state ϕ in it is called* in principle *Bayes accessible from another state $\psi \neq \phi$ if there exists a state space extension $L^1(X', \mathcal{S}', p')^\sharp$ of $L^1(X, \mathcal{S}, p)^\sharp$ such that the extension of ϕ from $L^1(X, \mathcal{S}, p)$ to $L^1(X', \mathcal{S}', p')$ is Bayes accessible from the extension of ψ from $L^1(X, \mathcal{S}, p)$ to $L^1(X', \mathcal{S}', p')$.*

DEFINITION 8.5. *The state space $L^1(X, \mathcal{S}, p)^\sharp$ is called* strongly *Bayes connectable if any state ϕ is in principle Bayes accessible from any other state ψ.*

PROBLEM 8.6. *Are state spaces strongly Bayes connectable*?

We do not know the answer to the above question.

§9. Summary and closing comments. Bayesian learning is a particular way of inferring unknown probabilities from known ones. The specificity of this kind of learning is that the inference is conditionalizing: the inferred probability measure is obtained by conditionalizing the known probability measure. We argued in this paper that conditionalizing should be carried out in terms of conditional expectations. We have seen that conditionalizing using this technique, which is standard in mathematics, includes both the elementary Bayes rule and Jeffrey conditionalization as special cases. We have shown that adopting this viewpoint leads naturally to regarding conditionalization as a two-place relation $\overset{\mathbb{E}}{\rightsquigarrow}$ in the state space determined by the reference probability measure representing the background subjective degrees of belief of a Bayesian Agent. The interpretation of $\psi \overset{\mathbb{E}}{\rightsquigarrow} \phi$ is that the Agent can learn the probabilities given by ϕ from the evidence represented by probabilities given by ψ; where "learning ϕ from ψ" means "conditionalizing ψ one obtains ϕ". Finding out the properties of the relation $\overset{\mathbb{E}}{\rightsquigarrow}$ amounts to characterizing Bayesian learning in its abstract, general form.

We have proved that the Bayes accessibility relation $\overset{\mathbb{E}}{\rightsquigarrow}$ is reflexive, antisymmetric, and nontransitive. We also have investigated the connectivity properties of state spaces with respect to the Bayes accessibility relation $\overset{\mathbb{E}}{\rightsquigarrow}$. We have shown that state spaces are typically not weakly Bayes connected. That is to say, we proved that, given a measure representing the background degrees of belief of a Bayesian Agent, there exist states (probability measures) that cannot be learned by the Agent from any evidence the Agent is capable of formulating within the confines of a given probability measure space. Failure of weak Bayes connectedness seems to pose a serious challenge for Bayesian learning: The existence of Bayes inaccessible states (we have proved that there exist an uncountably infinite number of such states in the typical cases) means that an Agent's background measure might prohibit the Agent from learning the "true" probability measure. By saying that a probability (measure) is "true" we only mean here that the probability measure is the one the Agent wishes to learn. The measure to be learned can be "true" in the sense of reflecting some objectively determined distribution (frequencies, ratios, etc.). But it also can be the degree of belief of another Agent. Nothing of what we presented in this paper depends on the specific nature of "true". Note also that an assignment of truth values *true* and *false* to elements of the Boolean algebra $\mathcal{S}$ that is in harmony with classical propositional logic is a specific probability measure, so learning such truth values is a specific case of learning a probability measure. If the true probability measure happens to be one of the Bayes inaccessible ones, the Agent cannot learn it by conditionalizing. Metaphorically speaking: The Agent's background degrees of belief creates "Bayes blind spots" for the Agent. Thus the Agent's background knowledge proves to be crucial from the perspective of what the Agent can in principle learn from possible evidence. In particular, the state spaces of standard probability measure spaces are not weakly Bayes connected. This is a very large class that includes most applications. In these probability theories the true probability measure to be learned might remain inaccessible for the Bayesian Agent. We also have seen that lack of weak Bayes connectedness cannot be "cured" by weakening weak Bayes connectedness to finite weak Bayes connectedness.

A natural question that arises is how "large" the Bayes Blind Spot is. There is no unique answer to this question, in part because there is no unique measure to gauge the size of a set: In addition to asking what its cardinality is, one can enquire about its topological size with respect to some topology in the state space and also about its size with respect to some natural measure on the state space. The paper [16] investigates the problem of size of the Byes Blind Spot in the case where the Boolean algebra is finite. It turns out that the Bayes Blind Spot in this case is as large as possible both in cardinality, in the sense of topology and in the natural measure: The Bayes Blind Spot is a set of second Baire category and its complement, the set of Bayes accessible states is a measure zero set. The topological and measure theoretic sizes of the Bayes Blind Spot of standard probability spaces are not known.

A Bayesian Agent might try to overcome the epistemological difficulty posed by Bayes inaccessible states by widening the probabilistic framework in which Bayes inaccessible states are present. This strategy involves enlarging the σ-algebra of propositions stating features of the world, extending the background probability to the enlarged set, and looking for evidence about (probabilities of) some subset of the enlarged σ-algebra—all this in the hope of becoming able to Bayes-learn those probabilities in the broader framework that are inaccessible in a narrower probability theory. We showed that such a strategy is in principle viable: a Bayes inaccessible state becomes Bayes learnable from some state after a suitable embedding of the original probability space into a larger one. Thus, a Bayesian Agent has unlimited learning capacity if he is allowed to expand the propositional base of possible evidence. It is even possible to enlarge the probability space into one in which Bayes inaccessible states do not exist and thus every probability is Bayes learnable from some evidence: We showed that state spaces of large enough probability spaces are weakly Bayes connected. A Bayesian Agent can only do such an extension however if he is capable of comprehending a very large amount of propositions: The σ-algebra of the probability space which could be shown having a Bayes connected state space had cardinality larger than the cardinality of the set of Lebesgue measurable subsets of real numbers. Since the cardinality of the set of Lebesgue measurable sets itself is already beyond the continuum, one needs an extremely strong concept of Bayesian Agent to allow for this option. A Bayesian Agent with a more modest mental capacity has to be aware however that he is on an unended quest: for him in every probability space he is able to comprehend there exist probability statements that might be true but he only can learn them from evidence that can be gathered only by going beyond the framework in which the true probability is formulated. Whether the concept of a powerful Bayesian Agent is reasonable, and whether the notion of a modest Bayesian Agent is attractive, we do not wish to try to decide here.

§10. Acknowledgement. Research supported by the National Research, Development, and Innovation Office, K 115593 and K 100715. M. Rédei thanks the Munich Center for Mathematical Philosophy in the Ludwig-Maximilians University (Munich, Germany), with which he was affiliated as Visiting Fellow, and the Institute of Philosophy of the Hungarian Academy of Sciences, where he was an Honorary Research Fellow, while this paper was written.

BIBLIOGRAPHY

[1] Billingsley, P. (1995). *Probability and Measure* (third edition). New York, Chichester, Brisbane, Toronto, Singapore: John Wiley & Sons.

[2] Bogachev, V. I. (2007). *Measure Theory*, Vol. II. Berlin, Heidelberg, New York: Springer.

[3] Bovens, L. & Hartmann, S. (2004). *Bayesian Epistemology*. Oxford, UK: Oxford University Press.

[4] Diaconis, P. & Zabell, S. L. (1982). Updating subjective probability. *Journal of the American Statistical Association*, **77**, 822–830.

[5] Doob, J. (1996). The development of rigor in mathematical probability theory (1900–1950). *American Mathematical Monthly*, **103**(7), 586–595.

[6] Doob, J. L. (1953). *Stochastic Processes*. New York: John Wiley & Sons.

[7] Douglas, R. G. (1965). Contractive projections on an L_1-space. *Pacific Journal of Mathematics*, **15**(2), 443–462.

[8] Earman, J. (1992). *Bayes or Bust*? Cambridge, Massachusetts: MIT Press.

[9] Easwaran, K. (2008). The Foundations of Conditional Probability. Ph.D. Thesis, University of California at Berkeley.

[10] Easwaran, K. (2011). Bayesianism I: Introduction and arguments in favor. *Philosophy Compass*, **6**, 312–320.

[11] Easwaran, K. (2011) Bayesianism II: Applications and criticisms. *Philosophy Compass*, **6**, 321–332.

[12] Feller, W. (1966). *An Introduction to Probability Theory and its Applications, Second Edition*, Vol. 2. New York: Wiley. First edition: 1966.

[13] Fremlin, D. H. (2001). *Measure Theory*, Vol. 2. Colchester, England: Torres Fremlin.

[14] Gyenis, B. (2015). Bayes rules all. Submitted.

[15] Gyenis, Z., Hofer-Szabó, G., & Rédei, M. (2016). Conditioning using conditional expectations: The Borel-Kolmogorov Paradox. *Synthese*, forthcoming, online March 26, 2016, doi: 10.1007/s11229-016-1070-8.

[16] Gyenis, Z. & Rédei, M. (2016). The Bayes Blind Spot of a finite Bayesian Agent is a large set. Manuscript.

[17] Hájek, A. (2003). What conditional probability could not be. *Synthese*, **137**, 273–333.

[18] Halmos, P. (1950). *Measure Theory*. New York: D. Van Nostrand.

[19] Hartmann, S. & Sprenger, J. (2010). Bayesian epistemology. In Bernecker, S. and Pritchard, D., editors. *Routledge Companion to Epistemology*. London: Routledge, pp. 609–620.

[20] Howson, C. (1996). Bayesian rules of updating. *Erkenntnis*, **45**, 195–208.

[21] Howson, C. (2014). Finite additivity, another lottery paradox, and conditionalization. *Synthese*, **191**, 989–1012.

[22] Howson, C. & Franklin, A. (1994). Bayesian conditionalization and probability kinematics. *The British Journal for the Philosophy of Science*, **45**, 451–466.

[23] Howson, C. & Urbach, P. (1989). *Scientific Reasoning: The Bayesian Approach*. Illinois: Open Court. Second edition: 1993.

[24] Huttegger, S. M. (2015). Merging of opinions and probability kinematics. *The Review of Symbolic Logic*, **8**, 611–648.

[25] Jaynes, E. T. (2003). Principles and pathology of orthodox statistics. In Larry Bretthorst, G., editor. *Probability Theory. The Logic of Science*. Cambridge: Cambridge University Press, pp. 447–483.

[26] Jeffrey, R. C. (1965). *The Logic of Decision* (first edition). Chicago: The University of Chicago Press.

[27] Kadison, R. V. & Ringrose, J. R. (1986). *Fundamentals of the Theory of Operator Algebras*, Vols. I. and II. Orlando: Academic Press.

[28] Kalmbach, G. (1983). *Orthomodular Lattices*. London: Academic Press.

[29] Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer. English translation: Foundations of the Theory of Probability (Chelsea, New York, 1956).

[30] Loéve, M. (1963). *Probability Theory* (third edition). Toronto, London, Melbourne: D. Van Nostrand, Princeton.

[31] Marchand, J.-P. (1977). Relative coarse-graining. *Foundations of Physics*, **7**, 35–49.

[32] Marchand, J.-P. (1981). Statistical inference in quantum mechanics. In Gustafson, K. E. and Reinhard, W. P., editors. *Quantum Mechanics in Mathematics, Chemistry, and Physics*. New York: Plenum Press, pp. 73–81. Proceedings of a special session in mathematical physics organized as a part of the 774th meeting of the American Mathematical Society, held March 27–29, 1980, in Boulder, Colorado.

[33] Marchand, J.-P. (1982). Statistical inference in non-commutative probability. *Rendiconti del Seminario Matematico e Fisico di Milano*, **52**, 551–556.

[34] Myrvold, W. (2015). You can't always get what you want: Some considerations regarding conditional probabilities. *Erkenntnis*, **80**, 572.

[35] Petersen, K. (1989). *Ergodic Theory*. Cambridge: Cambridge University Press.

[36] Pfanzagl, J. (1967). Characterizations of conditional expectations. *The Annals of Mathematical Statistics*, **38**, 415–421.

[37] Rao, M. M. (2005). *Conditional Measures and Applications* (second revised and expanded edition). Boca Raton, London, New York, Singapore: Chapman & Hall/CRC.

[38] Rédei, M. (1998). *Quantum Logic in Algebraic Approach*. Fundamental Theories of Physics, Vol. 91. Dordrecht, The Netherlands: Kluwer Academic Publisher.

[39] Rescorla, M. (2015). Some epistemological ramifications of the Borel-Kolmogorov Paradox. *Synthese*, **192**(3), 735–767.

[40] Roman, S. (2005). *Field Theory* (second edition). Graduate Texts in Mathematics, Vol. 158. New York: Springer.

[41] Rosenthal, J. S. (2006). *A First Look at Rigorous Probability Theory*. Singapore: World Scientific.

[42] Rudin, W. (1987). *Real and Complex Analysis* (third edition). Singapore: McGraw-Hill.

[43] Villani, A. (1985). Another note on the inclusion $L^p(\mu) \subset L^q(\mu)$. *The American Mathematical Monthly*, **92**, 485–487.

[44] Wagner, C. (2002). Probability kinematics and commutativity. *Philosophy of Science*, **69**, 266–278.

[45] Weisberg, J. (2011). Varieties of Bayesianism. In Gabbay, D. M., Hartmann, S., and Woods, J., editors. *Inductive Logic*. Handbook of the History of Logic, Vol. 10. Oxford: North-Holland (Elsevier), pp. 477–551.

[46] Weisberg, J. (2015). You've come a long way, Bayesians. *Journal of Philosophical Logic*, **44**, 817–834.

[47] Williamson, J. (2010). *In Defence of Objective Bayesianism*. Oxford: Oxford University Press.

DEPARTMENT OF ALGEBRA
BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS
BUDAPEST, HUNGARY
E-mail: gyz@renyi.hu

DEPARTMENT OF PHILOSOPHY, LOGIC AND SCIENTIFIC METHOD
LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE
HOUGHTON STREET, LONDON WC2A 2AE, UK
E-mail: m.redei@lse.ac.uk