Lower and Upper Conditioning in Quantum Bayesian Theory*

Bart Jacobs
Institute for Computing and Information Sciences (iCIS),
Radboud University Nijmegen, The Netherlands.
bart@cs.ru.nl

Updating a probability distribution in the light of new evidence is a very basic operation in Bayesian probability theory. It is also known as state revision or simply as conditioning. This paper recalls how locally updating a joint state can equivalently be described via inference using the channel extracted from the state (via disintegration).

This paper also investigates the quantum analogues of conditioning, and in particular the analogues of this equivalence between updating a joint state and inference. The main finding is that in order to obtain a similar equivalence, we have to distinguish two forms of quantum conditioning, which we call lower and upper conditioning. They are known from the literature, but the common framework in which we describe them and the equivalence result are new.

1 Introduction

This paper is about quantum analogues of Bayesian reasoning. It works towards one main result, Theorem 3 below, which gives a relation between locally updating a joint state and Bayesian inference. This is a fundamental matter, which requires some preparation in order to set the scene.

We use the term 'classical' probability for the ordinary, non-quantum form. We often use the word 'state' for a probability distribution, both in the classical and the quantum case. Classical Bayesian probability is based on what is called Bayes' rule. It describes probabilities of events (evidence) in an updated state. In fact, there are two closely related rules, sometimes called 'product rule' en 'Bayes rule' (proper). Making this distinction is not so relevant in the classical case, but, as we shall see, it is very relevant in the quantum case.

The paper starts with the back-and-forth constructions between a joint state (distribution) on the one hand, and a channel with an initial state on the other. A channel is a categorical abstraction of a conditional probability. We shall describe this process in terms of pairing and disintegration, following [4]. This process has a logical dimension that relates locally updating a joint state ('crossover inference') and Bayesian inference via the associated channel, in a result called the Bayesian Inference Theorem (see Theorem 2 below). This result is already described in [4], but is repeated here in more concrete form, and illustrated with an example.

The second part of the paper is about analogues in the quantum world. The constructions back-and-forth between a joint state and a channel exist in the literature [20] and are adapted to the current context. What is new here is the quantum logical analogue of this back-and-forth process. It is shown that updating a state with new evidence, in the form of a predicate, splits in two operations, which we call 'lower' and 'upper' conditioning. Both forms exist already, but not as counterparts. We show that

*The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement nr. 320571.

Giulio Chiribella, Peter Selinger (Eds.): 15th International
Conference on Quantum Physics and Logic (QPL 2018).
EPTCS 287, 2019, pp. 225–238, doi:10.4204/EPTCS.287.13

the earlier mentioned product rule holds for lower conditioning, but Bayes' rule itself holds for upper conditioning. In classical probability, the 'lower' and 'upper' versions coincide.

In a next step, the main result of the paper (Theorem 3) shows how 'lower' updating a joint state can equivalently be done via Bayesian inference with 'upper' conditioning, using the channel that is extracted from the joint state. This puts lower and upper conditioning into perspective and unveils some fundamental aspects of a quantum Bayesian theory.

Finally, there are two separate points worth emphasising. First, several constructions in this paper are illustrated with concrete calculations, via the Python-based tool EfProb [5]; it works both for classical and quantum probability and uses a common language for both. Next, along the way we find a novel result about how disintegration introduces 'semi' higher order structure in discrete probability, see Subsection 3.2.

2 Basics of discrete classical probability

This section recalls the basics of (classical, finite) discrete probability and fixes notation. For more information we refer to [14]. A *distribution*, also called a *state*, on a set X is a function $\omega\colon X \to [0,1]$ with finite support $\mathrm{supp}(\omega) = \{x \in X \mid \omega(x) \neq 0\}$ and with $\sum_x \omega(x) = 1$. Such a distribution can also be written as formal convex sum $\omega = \sum_x \omega(x)|x\rangle$. We write $\mathscr{D}(X)$ for the set of such distributions. The mapping $X \mapsto \mathscr{D}(X)$ is a monad on the category of sets, called the distribution monad.

A *joint* state is a state on an n-ary product set. A binary state is thus a distribution $\tau \in \mathscr{D}(X_1 \times X_2)$. It has first and second *marginals*, written here as $\mathsf{M}_1(\tau) \in \mathscr{D}(X_1)$ and $\mathsf{M}_2(\tau) \in \mathscr{D}(X_2)$. These marginal states are defined in the standard way as $\mathsf{M}_1(\tau)(x_1) = \sum_{x_2} \tau(x_1,x_2)$ and $\mathsf{M}_2(\tau)(x_2) = \sum_{x_1} \tau(x_1,x_2)$. In the other direction, two states $\omega_i \in \mathscr{D}(X_i)$ can be combined to product state $\omega_1 \otimes \omega_2 \in \mathscr{D}(X_1 \times X_2)$ via $(\omega_1 \otimes \omega_2)(x_1,x_2) = \omega_1(x_1) \cdot \omega_2(x_2)$. Obviously, $\mathsf{M}_i(\omega_1 \otimes \omega_2) = \omega_i$.

A *channel* is a function of the form $c\colon X \to \mathscr{D}(Y)$, that is, a map $X \to Y$ in the Kleisli category $\mathscr{K}\ell(\mathscr{D})$ of the distribution monad $\mathscr{D}$. Such a channel c has a *Kleisli extension* function, or *state transformer*, $c \gg (-)\colon \mathscr{D}(X) \to \mathscr{D}(Y)$ given by $(c \gg \omega)(y) = \sum_x \omega(x) \cdot c(x)(y)$. For another channel $d\colon Y \to \mathscr{D}(Z)$ there is a composite channel $d \bullet c\colon X \to \mathscr{D}(Z)$ given by $(d \bullet c)(x) = d \gg c(x)$. Channels $c_i\colon X_i \to \mathscr{D}(Y_i)$ can be combined to a product channel $c_1 \otimes c_2\colon X_1 \times X_2 \to \mathscr{D}(Y_1 \times Y_2)$ by $(c_1 \otimes c_2)(x_1,x_2) = c_1(x_1) \otimes c_2(x_2)$.

A (fuzzy) predicate on a set X is a function $p\colon X \to [0,1]$. For another predicate $q \in [0,1]^X$ there is a (sequential) conjunction predicate $p \,\&\, q$ on X via $(p \,\&\, q)(x) = p(x) \cdot q(x)$. For two predicates $p_i \in [0,1]^{X_i}$ on different sets X_i we can form a parallel conjunction predicate $p_1 \otimes p_2 \in [0,1]^{X_1 \times X_2}$, given by $(p_1 \otimes p_2)(x_1,x_2) = p_1(x_1) \cdot p_2(x_2)$. There is always a truth channel $\mathbf{1} \in [0,1]^X$ given by $\mathbf{1}(x) = 1$.

For a state $\omega \in \mathscr{D}(X)$ and a predicate $p \in [0,1]^X$ on the same set X the validity $\omega \models p$ in $[0,1]$ is the expected value $\sum_x \omega(x) \cdot p(x)$. If this validity is non-zero, one can form a conditioned state $\omega|_p$ on X, given by $\omega|_p(x) = \frac{\omega(x) \cdot p(x)}{\omega \models p}$. This updated state $\omega|_p$ is called 'ω given p', and is commonly written as $\omega(-|p)$. It is easy to check to see that conditioning with truth does nothing: $\omega|_{\mathbf{1}} = \omega$.

Proposition 1. *Assuming the conditionings of the states below exist, we have the 'product' rule on the left, and the 'Bayesian' rule on the right:*

$$\omega|_p \models q \;=\; \frac{\omega \models p \,\&\, q}{\omega \models p} \qquad\qquad \omega|_p \models q \;=\; \frac{(\omega|_q \models p) \cdot (\omega \models q)}{\omega \models p}. \tag{1}$$

Moreover, successive conditioning can be reduced to a single conditioning, as on the left below, so that conditioning becomes commutative, as on the right:

$$(\omega|_p)|_q \;=\; \omega|_{p \& q} \qquad\qquad (\omega|_p)|_q \;=\; (\omega|_q)|_p. \tag{2}$$

Proof The first equation in (1) follows simply by unravelling the definitions. The second equation directly follows from the first one, using that conjunction & is commutative. Similarly one obtains the equations in (2). □

Each channel $c\colon X \to \mathscr{D}(Y)$ also gives rise to a *predicate transformer* function $c \ll (-)\colon [0,1]^Y \to [0,1]^X$, given by $(c \ll q)(x) = \sum_y c(x)(y) \cdot q(y)$. We can now relate validity $\models$ and state/predicate transformation ($\gg$ and $\ll$) via the following fundamental equality of validities:

$$(c \gg \omega) \models q \quad = \quad \omega \models (c \ll q). \tag{3}$$

3 Classical Bayesian nets and disintegration

A major rationale for using Bayesian networks [23, 3, 2, 19] is efficiency of representation: a joint probability distribution (state) on multiple sample spaces (domains) quickly becomes very large. Representing the same distribution in graphical form, as a 'Bayesian network' is often much more efficient. The directed graph structure is determined by conditional independence. Semantically, the directed arcs are given by channels, that is by stochastic matrices, or more abstractly by Kleisli morphisms for the distribution monad [9, 16, 17].

The essence of this semantical view on Bayesian network theory consists of two parts.

(I) The ability to move back-and-forth between a joint state and a graph (network) of channels. The difficult direction is extracting the various channels of the graph from a joint state. This is often called *disintegration* [4].

(II) Equivalence of inference via joint states and inference via associated channels. In general, inference (or, Bayesian learning) happens via conditioning (updating, revising) of states, in the light of evidence given by predicates. Inference involves the propagation of such conditioning via joint states and/or via channels, via the bank-and-forth connections in (I), both in a forward and backward direction (as in [16, 18]).

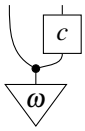
Point (I) is well-known, but point (II) is usually left implicit; it is however a crucial part of why efficient representation of (big) joint states as Bayesian network graphs can be used for Bayesian reasoning. In this section we briefly elaborate both points below, and illustrate them with an example.

Note that we do not claim that with these two points (I) and (II) we capture all essentials of Bayesian network theory: *e.g.*, we do not address the matter of *how* to turn a joint state into a graph, via conditional independence or via causality. This question has also be studied in a quantum setting, see *e.g.* [24].

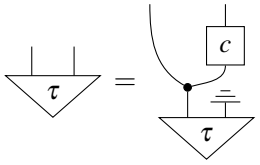
3.1 Disintegration

Abstractly, point (I) involves the correspondence between a joint state on the one hand, and a channel and a (single) state on the other hand. In one direction this is easy: given a state ω on X and a channel $c\colon X \to Y$ we can form a joint state on $X \times Y$, namely as: $\mathrm{pair}(\omega, c) \coloneqq \big((\mathrm{id} \otimes c) \bullet \Delta\big) \gg \omega$, where $\Delta\colon X \to X \times X$ is the copier channel with $\Delta(x) = 1|x,x\rangle$. This construction is drawn as a picture on the right (4), using the graphical language associated with monoidal categories. It will be used here only as illustration, hopefully in an intuitive self-explanatory manner. We refer to [25, 7, 4] for details.

$$\mathrm{pair}(\omega, c) \coloneqq \tag{4}$$



Going in the other direction, from a joint state to a channel is less trivial. It is called *disintegration e.g.* in [4]. It involves a joint state τ on X,Y from which a channel $c\colon X \to Y$ is extracted, in such a way that τ itself can be reconstructed from its first marginal $\mathsf{M}_1(\tau)$ and the channel c. Pictorially this marginal is represented by blocking its second wire via the ground symbol $\perp$. We write $\mathrm{extr}(\tau)$ for this extracted channel c. Then we can write Equation (5) as $\tau = \mathrm{pair}\big(\mathsf{M}_1(\tau), \mathrm{extr}(\tau)\big)$.



(5)

Lemma 2. *Extracted channels* $\mathrm{extr}(\tau)$ *exist and are unique in classical discrete probability, for joint states* τ *whose first marginal has full support.*

Proof First, a state $\omega \in \mathscr{D}(X)$ and a channel $c\colon X \to \mathscr{D}(Y)$ yield a joint state $\mathrm{pair}(\omega,c) \in \mathscr{D}(X \times Y)$, namely, as described in (4) above:

$$\mathrm{pair}(\omega,c)(x,y) \quad = \quad \big(((\mathrm{id} \otimes c) \bullet \Delta) \gg \omega\big)(x,y) \quad = \quad \omega(x) \cdot c(x)(y).$$

In the other direction, let $\tau \in \mathscr{D}(X \times Y)$ be a joint state whose first marginal $\mathsf{M}_1(\tau) \in \mathscr{D}(X)$ has full support. The latter means that its support is the whole of X, so that: $\mathsf{M}_1(\tau)(x) \neq 0$ for each x, or, equivalently, $\forall x.\, \exists y.\, \tau(x,y) \neq 0$. We can now define a channel $\mathrm{extr}(\tau)\colon X \to \mathscr{D}(Y)$ by:

$$\mathrm{extr}(\tau)(x)(y) \quad = \quad \frac{\tau(x,y)}{\mathsf{M}_1(\tau)(x)} \quad = \quad \frac{\tau(x,y)}{\sum_z \tau(x,z)}. \qquad \square$$

For a more systematic, diagrammatic description of disintegration, also for continuous probability, we refer to [4]. Here we only need it for discrete probability, as a preparation for the quantum case.

3.2 Excursion on disintegration and semi-exponentials

We conclude this part on disintegration with a novel observation. It is interesting in itself, but it does not play a role in the sequel. It shows that disintegration gives rise to higher order 'semi-exponential' structure, originally introduced in [11]. Recall that a categorical description of a (proper) exponential in a cartesian closed category involves exponent objects Y^X with an evaluation map $\mathrm{ev}\colon Y^X \times X \to Y$ such that for each map $f\colon Z \times X \to Y$ there is an abstraction map $\Lambda(f)\colon Z \to Y^X$. These ev and Λ should satisfy:

$$\mathrm{ev} \circ (\Lambda(f) \times \mathrm{id}) \quad = \quad f \qquad\qquad \Lambda(f \circ (g \times \mathrm{id})) \quad = \quad \Lambda(f) \circ g \qquad\qquad \Lambda(\mathrm{ev}) \quad = \quad \mathrm{id}. \tag{6}$$

The last two equations ensure that $\Lambda(f)$ is the unique map h with $\mathrm{ev} \circ (h \times \mathrm{id}) = f$, since: $h = \mathrm{id} \circ h = \Lambda(\mathrm{ev}) \circ h = \Lambda(\mathrm{ev} \circ (h \times \mathrm{id})) = \Lambda(f)$.

For a *semi-exponential*, the last equation in (6) need not hold. A semi-exponential is thus more than a 'weak' exponential (only the first equation) since it also satisfies naturality (the second equation). In the language of the λ-calculus, having 'semi-exponentials' means that one has a β-equation, but not an η-equation, see [11] or [12] for more details.

Theorem 1. *Let* $\mathscr{K}\ell_{\mathsf{f}}(\mathscr{D})$ *be the subcategory of the Kleisli category* $\mathscr{K}\ell(\mathscr{D})$ *of the distribution monad on* **Sets** *with only* finite *sets as objects. This category* $\mathscr{K}\ell_{\mathsf{f}}(\mathscr{D})$ *is symmetric monoidal 'semi' closed: it has semi-exponentials, which are semi-right adjoint to the (standard) tensor product.*

Proof It is well-known that cartesian products $\times$ on sets and parallel product $\otimes$ on Kleisli maps (channels) makes the category $\mathscr{K}\ell(\mathscr{D})$, and also $\mathscr{K}\ell_{\mathsf{f}}(\mathscr{D})$, symmetric monoidal closed. We sketch how semi-exponentials $\multimap$ are obtained via joint states whose first marginal has full support:

$$X \multimap Y \quad := \quad \{\tau \in \mathscr{D}(X \times Y) \mid \operatorname{supp}\big(\mathsf{M}_1(\tau)\big) = X\}.$$

This definition assumes that X is not the empty set. In that case we can set $\emptyset \multimap Y = 1$, the singleton set, since $Z \times \emptyset \cong \emptyset$ so that there is a trivial correspondence between maps $Z \times \emptyset \to \mathscr{D}(Y)$ and maps $Z \to \mathscr{D}(1) = 1$.

We define an evaluation channel $\mathrm{ev}\colon (X \multimap Y) \times X \to \mathscr{D}(Y)$ via disintegration:

$$\mathrm{ev}(\tau, x)(y) \quad := \quad \mathrm{extr}(\tau)(x)(y) \quad = \quad \frac{\tau(x,y)}{\sum_z \tau(x,z)}.$$

For abstraction, let $f\colon Z \times X \to \mathscr{D}(Y)$ be given. We define $\Lambda(f)\colon Z \to \mathscr{D}(X \multimap Y)$ as:

$$\Lambda(f)(z) \quad := \quad 1\big|\,\tau\big\rangle \qquad \text{where} \qquad \tau(x,y) \quad = \quad \frac{f(z,x)(y)}{\#X},$$

and where $\#X$ is the number of elements in X. Here we construct a joint state $\tau = \mathrm{pair}(\frac{1}{\#}, f(z,-))$ as in the beginning of this section, from the uniform distribution $\frac{1}{\#}$ on X and the channel $f(z,-)\colon X \to \mathscr{D}(Y)$. We need to check that $\Lambda(f)$ is well-defined, in particular that each first marginal of $\Lambda(f)(z)(*) \in \mathscr{D}(X \times Y)$ has full support:

$$\mathsf{M}_1(\Lambda(f)(z)(*))(x) \quad = \quad \textstyle\sum_y \Lambda(f)(z)(*)(x,y) \quad = \quad \sum_y \frac{f(z,x)(y)}{\#X} \quad = \quad \frac{1}{\#X} \quad \neq \quad 0.$$

It is easy to check that the first two equations from (6) hold. □

3.3 Bayesian inference and disintegration

We now turn to the second point (II) from the very beginning of this section, about Bayesian inference, especially in relation to the passage back-and-forth between joint states and channels via pairing and extraction, as just described.

It may happen that a joint state $\tau \in \mathscr{D}(X \times Y)$ is equal to the product of its two marginals, *i.e.* $\tau = \mathsf{M}_1(\tau) \otimes \mathsf{M}_2(\tau)$. The state τ is then called *non-entwined*. The more common case is that a joint state is entwinted, and its marginal components are correlated. If we then update in one component, we see a change in the other component. This is called *crossover influence* in [17, 18].

The essence of the point (II), in the beginning of this section, about inference and disintegration is that for a joint state τ, this crossover influence can be propagated through the channel c that is extracted from the state τ via disintegration. This is expressed in the next result, called the *Bayesian Inference Theorem*.

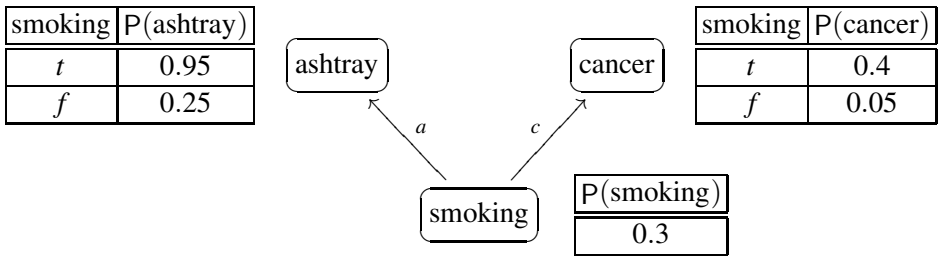
Theorem 2. *Let $\tau \in \mathscr{D}(X \times Y)$ be a joint state, and $c = \mathrm{extr}(\tau)\colon X \to \mathscr{D}(Y)$ the extracted channel obtained via disintegration — as described in Subsection 3.1. For predicates $p \in [0,1]^X$ and $q \in [0,1]^Y$ we then have:*

$$\mathsf{M}_2\big(\tau|_{p \otimes \mathbf{1}}\big) \quad = \quad c \gg \big(\mathsf{M}_1(\tau)\big|_p\big) \qquad \textit{and} \qquad \mathsf{M}_1\big(\tau|_{\mathbf{1} \otimes q}\big) \quad = \quad \mathsf{M}_1(\tau)\big|_{c \ll q}. \tag{7}$$

The first equation describes crossover inference on the left-hand-side as *forward* inference on the right: first update and then do state transformation $\gg$. The second equation in (7) describes crossover inference in the other component as *backward* inference: first do predicate transformation $\ll$ and then update. The terminology of 'forward' and 'backward' inference comes from [16], see also [18]. An abstract graphical proof of the equations (7) is given in [4]. But it is not hard to prove these equations concretely, by unwrapping the definitions.

3.4 An illustration of inference in a classical Bayesian network

We consider the relation between smoking and the presence of ashtrays and (lung) cancer, in the following simple Bayesian network.



Thus, 95% of people who smoke have an ashtray in their home, and 25% of the non-smokers too. On the right we see that in this situation a smoker has 40% chance of developing cancer, whereas a non-smoker only has 5% chance.

The question we want to address is: what is the influence of the presence or absence of an ashtray on the probability of developing cancer? Here the presence/absence of the ashtray is the 'evidence', whose influence is propagated through the network. We shall describe the outcome using the EfProb tool [5], concentrating on evidence propagation, and not so much on the precise representation of the above network, using channels `a` and `c` associated with the conditional probability tables.

We first consider the prior probabilities of smoking, ashtray, and cancer:

```
>>> smoking
0.3|t> + 0.7|f>
>>> a >> smoking
0.46|t> + 0.54|f>
>>> c >> smoking
0.155|t> + 0.845|f>
```

The network gives rise to a joint state, by tupling the ashtray, identity and cancer channels, and applying them to the `smoking` state. We can then obtain the above three prior probabilities alternatively via three marginalisations of this joint state, namely as first, second, third marginals, by using in EfProb the corresponding masks `[1,0,0]`, `[0,1,0]`, `[0,0,1]` after the marginalisation sign `%`.

```
>>> joint = (a @ idn(bnd) @ c) * copy(bnd,3) >> smoking
>>> joint
0.114|t,t,t> + 0.171|t,t,f> + 0.00875|t,f,t> + 0.166|t,f,f>
 + 0.006|f,t,t> + 0.009|f,t,f> + 0.0263|f,f,t> + 0.499|f,f,f>
>>> joint % [1,0,0]
0.46|t> + 0.54|f>
>>> joint % [0,1,0]
0.3|t> + 0.7|f>
>>> joint % [0,0,1]
0.155|t> + 0.845|f>
```

We now wish to infer the (adapted) cancer probability when we have evidence of ashtrays. We shall do this in two ways, first via crossover inference using the above joint state. The ashtray evidence `tt` needs to be extended (weakened) to a predicate with the same domain as the joint state. In the Equations (7) this is written as: $p \otimes \mathbf{1}$, but in EfProb it is: `tt @ truth(bnd) @ truth(bnd)`. We first use this predicate

for updating the joint state, written as / in EfProb, and then we marginalise to obtain the third 'cancer' component that we are interested in:

```
>>> (joint / (tt @ truth(bnd) @ truth(bnd))) % [0,0,1]
0.267|t> + 0.733|f>
```

Alternatively we can compute this posterior cancer probability by following the graph structure. The ashtray evidence `tt` is now first turned into predicate `a << tt` on the state `smoking`. After updating this state, we transform it to an updated cancer probability, via state transformation `>>`. We can do this down-and-up propagation in one go:

```
>>> c >> (smoking / (a << tt))
0.267|t> + 0.733|f>
```

The fact that we get the same distribution is an instance of the equations (7). As expected, in presence of ashtrays the probability of cancer is higher.

Aside: clearly, ashtrays *influence* (the probability of) cancer, but they are not the *cause*; in the graph this influence happens via a common ancestor, namely smoking, working statistically as 'confounder', and as the actual cause of cancer.

4 Towards quantum Bayesian theory

The main aim of this paper is to investigate quantum analogues of the Bayesian Inference Theorem 2, from the conviction that any adequate quantum Bayesian network theory should address these points (I) and (II) from the beginning of Section 3 in a satisfactory manner. Point (I) has received ample attention in quantum theory, see for instance [20, 8, 21, 1]. But Point (II) involving quantum conditioning has not really been studied this explicitly. Our main result is that one can also describe quantum conditioning consistently, both on joint states and via channels, as in Equations (7), but this requires in the quantum case that one distinguishes *two forms of conditioning*, which we shall call *lower* and *upper* conditioning, written as $\sigma|_p$ and $\sigma|^p$ respectively[1]. Classically these two forms of conditioning coincide, but the quantum world is more subtle — as usual. Lower conditioning has appeared in effectus theory [6] and upper conditioning in the approach of [20]. Here they are clearly distinguished for the first time, and used jointly to capture quantum inference and propagation of evidence. Interestingly, what is commonly called Bayes' rule holds for upper conditioning, but not for lower conditioning, for which we "only" have the product rule.

First we introduce the basics about states and predicates in the quantum world. We shall do so for finite-dimensional quantum theory, using the formalism of Hilbert spaces.

4.1 Basics of quantum probability

Let $\mathscr{H}$ be a finite-dimensional complex Hilbert space. A *state* σ of $\mathscr{H}$ is a positive operator on $\mathscr{H}$ with trace one. That is, σ is linear function $\sigma\colon \mathscr{H} \to \mathscr{H}$ satisfying $\sigma \geq 0$ and $\mathrm{tr}(\sigma) = 1$. A state is often called a *density matrix*. The canonical way to define a state is to start from a vector $|v\rangle \in \mathscr{H}$ with norm 1, and consider the operator $|v\rangle\langle v|\colon \mathscr{H} \to \mathscr{H}$. It sends any element $|w\rangle \in \mathscr{H}$ to the vector $\langle v|w\rangle \cdot |v\rangle$. An arbitrary state is a convex combination of such vector states $|v\rangle\langle v|$. A *joint* state τ on two Hilbert space $\mathscr{H}$ and $\mathscr{K}$ is a state on the tensor product $\mathscr{H} \otimes \mathscr{K}$.

[1] The terminology 'lower' and 'upper' is simply determined by the position of the predicate p, low in $\sigma|_p$ and up in $\sigma|^p$.

A *predicate*, also called an *effect*, is a positive operator p on $\mathscr{H}$ below the identity: $0 \leq p \leq \mathrm{id}$. The identity id is given by the identity/unit matrix, and corresponds to the *truth* predicate, often written as $\mathbf{1}$. For each predicate p there is an orthosupplement, written as $p^\perp$, playing the role of negation. It is defined by $p^\perp = \mathrm{id} - p$, and satisfies: $p^{\perp\perp} = p$ and $p + p^\perp = \mathbf{1}$.

The most interesting logical operation on quantum predicates is *sequential conjunction* $\&$. It is defined via the square root operation on predicates, as:

$$p \,\&\, q \quad = \quad \sqrt{p}\, q \sqrt{p}. \tag{8}$$

We pronounce $\&$ as 'and-then', and read it as: after p with its side-effect, the predicate q holds. This operation $\&$ has been studied in [10], and re-emerged in effectus theory [13, 6]. The square root of the matrix p exists since p is positive. It is computed via diagonalisation $\sqrt{p} = U\sqrt{D}U^{-1}$, where $p = UDU^{-1}$, in which $\sqrt{D}$ is obtained from the diagonal matrix D by taking the square roots of the (positive) eigenvalues on the diagonal.

States σ and predicates p of the same Hilbert space $\mathscr{H}$ can be combined in *validity*, defined as:

$$\sigma \models p \quad := \quad \mathrm{tr}(\sigma\, p) \;\in\; [0,1]. \tag{9}$$

This standard definition is also known as the Born rule.

Remark 1. *There is a standard way to embed classical probability into quantum probability. Suppose we have classical state ω and predicate p on a finite set $X = \{x_1, \ldots, x_n\}$ with n elements. Then we consider the Hilbert space $\mathbb{C}^n$ with standard basis given by vectors $|i\rangle$ with an 1 on the i-th position and zeros elsewhere. We write $\widehat{\omega} = \sum_i \omega(x_i)|i\rangle\langle i|$ for the 'diagonal' quantum state on $\mathbb{C}^n$. By construction it is positive and has trace $\sum_i \omega(x_i) = 1$.*

Similarly, a classical predicate $p \in [0,1]^X$ gives a quantum predicate $\widehat{p}$ on $\mathbb{C}^n$ via $\widehat{p} = \sum_i p(x_i)|i\rangle\langle i|$. By construction, $0 \leq p \leq \mathrm{id}$. It is easy to see that the classical and quantum validities coincide:

$$\omega \models p \quad = \quad \sum_i \omega(i) \cdot p(i) \quad = \quad \mathrm{tr}\big(\widehat{\omega}\,\widehat{p}\big) \quad = \quad \widehat{\omega} \models \widehat{p}.$$

The mapping $\widehat{(\,\cdot\,)}$ preserves the logical structure on predicates, including sequential conjunction $\&$.

Remark 2. *In both classical and quantum probability, as described here, a state is also a predicate. This is peculiar. When one moves to a higher level of abstraction, this is no longer the case — for instance by using von Neumann algebras instead of Hilbert spaces, or by using continuous probability distributions on measurable spaces instead of discrete distributions on sets. In the next section we sometimes 'convert' a state into a predicate, but we shall make explicit when we do so. A more abstract approach is possible, using the duality between states and effects, see also Remark 4.*

4.2 Two forms of quantum conditioning

This subsection introduces two forms of quantum conditioning of a state by a predicate, called 'lower' and 'upper' conditioning, and describes their basic properties.

Definition 1. Let σ be a state, and p a predicate, on the same Hilbert space, for which the validity $\sigma \models p$ is non-zero. We shall use the following terminology, notation and definition for two forms of conditioning:

$$\text{lower:} \quad \sigma|_p \;:=\; \frac{\sqrt{p}\,\sigma\sqrt{p}}{\sigma \models p} \qquad\qquad \text{upper:} \quad \sigma|^p \;:=\; \frac{\sqrt{\sigma}\,p\sqrt{\sigma}}{\sigma \models p}.$$

It is easy to see that both $\sigma|_p$ and $\sigma|^p$ are states again — using the familiar 'rotation' property of traces: $\mathrm{tr}(AB) = \mathrm{tr}(BA)$. Lower conditioning $\sigma|_p$ arises in effectus theory, whereas upper conditioning $\sigma|^p$ comes from [20]. We first observe that this difference between 'lower' and 'upper' does not exist classically.

Lemma 3. *For classical (non-quantum) states and predicates, lower and upper conditioning coincide with classical conditioning. To express this more precisely we use the notation $\widehat{(\,\cdot\,)}$ from Remark 1 to translate from classical to quantum: for a classical state ω and predicate p,*

$$\widehat{\omega}|_{\widehat{p}} \;=\; \omega|_p \;=\; \widehat{\omega}|^{\widehat{p}}.$$

Proof Diagonal matrices commute, so that $\sqrt{\widehat{p}}\,\widehat{\omega}\,\sqrt{\widehat{p}} = \widehat{\omega}\,\widehat{p} = \sqrt{\widehat{\omega}}\,\widehat{p}\,\sqrt{\widehat{\omega}}$. □

A second observation is about truth $\mathbf{1}$ and sequential conjunction &. Both lower and upper conditioning with truth $\mathbf{1}$ does nothing, like in the classical case, but successive conditioning cannot be reduced to single conditioning, like in the first equation in (2), in Proposition 1. In addition, the order in quantum conditioning matters, just like the order of priming in psychology matters [15].

Remark 3. *We have $\sigma|_{\mathbf{1}} = \sigma$ and $\sigma|^{\mathbf{1}} = \sigma$, but in general successive quantum conditionings cannot be reduced to a single conditioning via sequential conjunction:*

$$(\sigma|_p)|_q \;\neq\; \sigma|_{p\&q} \qquad \textit{and also} \qquad (\sigma|^p)|^q \;\neq\; \sigma|^{p\&q}.$$

Similarly, in general, quantum conditionings do not commute:

$$(\sigma|_p)|_q \;\neq\; (\sigma|_q)|_p \qquad \textit{and} \qquad (\sigma|^p)|^q \;\neq\; (\sigma|^q)|^p.$$

Interestingly, the two classical equations (1) in Proposition 1 hold separately for the two kinds of quantum conditioning.

Proposition 4. *The 'product' rule holds for lower conditioning and Bayes' rule holds for upper conditioning:*

$$\sigma|_p \models q \;=\; \frac{\sigma \models p \,\&\, q}{\sigma \models p} \qquad\qquad \sigma|^p \models q \;=\; \frac{(\sigma|^q \models p)\cdot(\sigma \models q)}{\sigma \models p}. \tag{10}$$

Proof We simply go through the computations:

$$\begin{aligned}
\sigma|_p \models q = \mathrm{tr}\big(\sigma|_p\, q\big) &= \mathrm{tr}\big(\tfrac{\sqrt{p}\,\sigma\sqrt{p}}{\sigma\models p}\, q\big) = \tfrac{1}{\sigma\models p}\cdot \mathrm{tr}\big(\sigma\sqrt{p}\,q\sqrt{p}\big) = \tfrac{\sigma\models p\&q}{\sigma\models p} \\
\sigma|^p \models q = \mathrm{tr}\big(\sigma|^p\, q\big) &= \mathrm{tr}\big(\tfrac{\sqrt{\sigma}\,p\sqrt{\sigma}}{\sigma\models p}\, q\big) = \tfrac{\sigma\models q}{\sigma\models p}\cdot \mathrm{tr}\big(p\tfrac{\sqrt{\sigma}\,q\sqrt{\sigma}}{\sigma\models q}\big) = \tfrac{(\sigma|^q\models p)\cdot(\sigma\models q)}{\sigma\models p}.
\end{aligned}$$ □

5 Quantum channels

In order to express the quantum analogues of the equations in Theorem 2 we need the notion of 'channel' in a quantum setting. It exists, and is alternatively often called a quantum operation, see *e.g.* [22]. There are several variations possible in the requirements, such as just positive or complete positive, unitary or subunitary, normal or not. These variations are not essential for what follows.

For a finite-dimensional Hilbert space $\mathscr{H}$ be write $\mathscr{B}(\mathscr{H})$ for the set of linear maps $A\colon \mathscr{H} \to \mathscr{H}$. Because $\mathscr{H}$ has finite dimension, such A are automatically bounded, or equivalently, continuous. The

set of operators $\mathscr{B}(\mathscr{H})$ is in fact a Hilbert space itself, with Hilbert-Schmidt inner product $\langle A|B\rangle_{\mathrm{HS}} = \mathrm{tr}(A^\dagger B)$, where $A^\dagger$ is the conjugate transpose of A, as matrix. Moreover, there are canonical isomorphisms $\mathscr{B}(\mathscr{H}\otimes\mathscr{K}) \cong \mathscr{B}(\mathscr{H})\otimes\mathscr{B}(\mathscr{K})$ and $\mathscr{B}(\mathbb{C}) \cong \mathbb{C}$.

If $\mathscr{K}$ is another finite-dimensional Hilbert space, then a *CP-map* $\mathscr{H} \to \mathscr{K}$ is a completely positive linear map $c\colon \mathscr{B}(\mathscr{K}) \to \mathscr{B}(\mathscr{H})$. Notice the change of direction. This CP-map c is called a *channel* if it preserves the unit/identity matrix: $c(\mathrm{id}) = \mathrm{id}$. It may be called *subchannel* if $c(\mathrm{id}) \leq \mathrm{id}$. Each CP-map $c\colon \mathscr{B}(\mathscr{K}) \to \mathscr{B}(\mathscr{H})$ has a 'dagger', written as $c^\#\colon \mathscr{B}(\mathscr{H}) \to \mathscr{B}(\mathscr{K})$, so that $\langle c(A)|B\rangle_{\mathrm{HS}} = \langle A|c^\#(B)\rangle_{\mathrm{HS}}$, that is, $\mathrm{tr}(c(A)^\dagger B) = \mathrm{tr}(A^\dagger c^\#(B))$.

For a channel $c\colon \mathscr{H} \to \mathscr{K}$ and a predicate (effect) q on $\mathscr{K}$ we define predicate transformation via function application $c \ll q := c(q)$. Similarly, for a state σ on $\mathscr{H}$ we define state transformation via the dagger of the channel, as: $c \gg \sigma := c^\#(\sigma)$. Then, using that positive operators are self-adjoint, we get the same relation (3) between validity and state/predicate transformation as in the classical case:

$$\begin{aligned} c \gg \sigma \models q \;=\; \mathrm{tr}\big(c^\#(s)\,q\big) \;&=\; \mathrm{tr}\big(c^\#(s)\,q^\dagger\big) \\ &= \mathrm{tr}\big(s\,c(q)^\dagger\big) \;=\; \mathrm{tr}\big(s\,c(q)\big) \;=\; s \models c \ll q. \end{aligned} \tag{11}$$

Definition 2. Let p be a (quantum) predicate on Hilbert space $\mathscr{H}$. It gives rise to a subchannel $\mathrm{asrt}_p\colon \mathscr{H} \to \mathscr{H}$ defined by:

$$\mathrm{asrt}_p(A) \;:=\; \sqrt{p}\,A\,\sqrt{p}.$$

This assert map asrt_p plays a fundamental role in effectus theory, see [13, 6], for instance because it allows us to define sequential conjunction (8) via predicate transformation as $p \mathbin{\&} q = \mathrm{asrt}_p \ll q$.

Remark 4. *States/predicates on $\mathscr{H}$ are special instances of CP-maps $\mathbb{C} \to \mathscr{H}$, resp. $\mathscr{H} \to \mathbb{C}$. If we consider them as such channels, we can take their dagger $(-)^\#$. Then we can relate upper and lower conditioning via an exchange, namely as: $\sigma|^p = p^\#|_{\sigma^\#}$. This re-formulation may be useful in a more general setting.*

5.1 Representation of quantum channels

As mentioned, a channel $c\colon \mathscr{H} \to \mathscr{K}$ is a (completely positive) linear function $\mathscr{B}(\mathscr{K}) \to \mathscr{B}(\mathscr{H})$ between spaces of operators. Let's assume $\mathscr{H},\mathscr{K}$ have dimensions n,m, respectively. The space of operators $\mathscr{B}(\mathscr{K})$ then has dimension $m \times m$, so that the channel c is determined by its values on the $m \times m$ base vectors $|i\rangle\langle j|$ of $\mathscr{B}(\mathscr{K})$. Thus, the channel c is determined by $m \times m$ matrices of size $n \times n$, as in:

$$\left(\begin{array}{ccc} \boxed{n\times n} & \cdots & \boxed{n\times n} \\ \vdots & & \vdots \\ \boxed{n\times n} & \cdots & \boxed{n\times n} \end{array}\right) \begin{array}{l} \uparrow \\ m \\ \downarrow \end{array} \qquad \leftarrow m \rightarrow \tag{12}$$

The matrix entries of the channel c will be written via double indexing, as $c_{k\ell,ij}$ for $1 \leq k,\ell \leq m$ and $1 \leq i,j \leq n$.

This matrix representation of a quantum channel is used in EfProb. It is convenient, for instance because parallel composition $\otimes$ of channels can simply be done by Kronecker multiplication of their (outer) matrices (12). We briefly describe how predicate and state transformation works.

Let q be a predicate on $\mathscr{K}$, represented as a $m \times m$ matrix. Predicate transformation $c \ll q$ is done simply by linear extension. It yields an $n \times n$ matrix, forming a predicate on $\mathscr{H}$, via:

$$c \ll q \quad := \quad \sum_{k,\ell} q_{k\ell} \cdot c_{k\ell}. \tag{13}$$

In the other direction we do state transformation essentially via the dagger $c^\#$ of the channel c. Explicitly, it works as follows. Let σ be a state of $\mathscr{H}$, represented by a $n \times n$ matrix. Then we obtain the transformed state $c \gg \sigma$ as an $m \times m$ matrix given by computing traces:

$$\big(c \gg \sigma\big)_{k\ell} \quad := \quad \mathrm{tr}(c_{\ell k}\, \sigma). \tag{14}$$

Notice the change of order of indices: at position (k,ℓ) of $c \gg \sigma$ we use the inner matrix $c_{\ell k}$ from (12). The reason is the implicit use of the Hilbert-Schmidt inner product, given by $\langle A|B\rangle_{\mathrm{HS}} = \mathrm{tr}(A^\dagger \cdot B)$, where the dagger involves a conjugate transpose.

5.2 Quantum pairing and extraction

The pairing of a classical state and a channel in (4) involves a copier $\curlyvee$. It does not exist in general in a quantum setting because of the 'no-cloning' theorem. But we do have 'cup' states $\cup$ with maximal entanglement. They are basis dependent: given a finite-dimensional Hilbert space $\mathscr{H}$ with orthonormal basis $\big(|i\rangle\big)$ of size n, we can for a state $\cup$ of $\mathscr{H} \otimes \mathscr{H}$ as $\cup = \frac{1}{n}\sum_{i,j} |ii\rangle\langle jj|$. Similarly, there is 'cap' predicate $\cap$. The quantum pairing and extraction operations that we describe in this subsection are due to [20]. But the more abstract description in terms of cups and caps does not occur there. These operations depend on a choice of basis.

Given a state σ of $\mathscr{H}$ and a channel $c\colon \mathscr{H} \to \mathscr{K}$ we can thus form a joint state of $\mathscr{H} \otimes \mathscr{K}$ via the 'cup' state $\cup$ of $\mathscr{H} \otimes \mathscr{H}$. Then we can define a pair state of $\mathscr{H} \otimes \mathscr{K}$ via state transformation $\gg$ as in:

$$\mathrm{pair}(\sigma, c) \quad := \quad \big(\mathrm{asrt}_{\sigma^T} \otimes c\big) \gg \cup \qquad \text{that is} \qquad \langle ik|\mathrm{pair}(\sigma,c)|j\ell\rangle \quad = \quad \overline{\big(\sqrt{\sigma} c_{k\ell} \sqrt{\sigma}\big)_{ij}}. \tag{15}$$

In the other direction, given a joint state τ of $\mathscr{H} \otimes \mathscr{K}$ we write $\mathrm{proj}(\tau)$ for the transpose of its first marginal, so:

$$\mathrm{proj}(\tau) \quad := \quad \mathsf{M}_1(\tau)^T \qquad \text{where} \qquad \langle i|\mathsf{M}_1(\tau)|j\rangle \quad := \quad \sum_k \langle ik|\tau|jk\rangle. \tag{16}$$

We extract a channel $\mathrm{extr}(\tau)\colon \mathscr{H} \to \mathscr{K}$ from τ in the manner defined in [20]:

$$\mathrm{extr}(\tau)_{k\ell} \quad := \quad \sum_{i,j} \overline{\langle ik|\tau|j\ell\rangle} \cdot \big(\sqrt{\mathrm{proj}(\tau)^{-1}}|i\rangle\langle j|\sqrt{\mathrm{proj}(\tau)^{-1}}\big). \tag{17}$$

The next result is the analogue of Lemma 2 about disintegration for classical discrete probability.

Proposition 5 (After [20])**.** *A quandum state σ and channel c, with matching types, can be recovered from their pair, defined in* (15)*, via projection* (16) *and extraction 17*

$$\mathrm{proj}(\mathrm{pair}(\sigma,c)) \quad = \quad \sigma \qquad \textit{and} \qquad \mathrm{extr}\big(\mathrm{pair}(\sigma,c)\big) \quad = \quad c.$$

Similarly, a joint state τ for which the transpose of its first marginal $\mathrm{proj}(\tau)$*, as defined above, is invertible can be recovered as a pair, as on the left below. In addition, τ's second marginal can be obtained via state transformation, as on the right:*

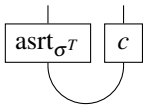
$$\tau \quad = \quad \mathrm{pair}(\mathrm{proj}(\tau), \mathrm{extr}(\tau)) \qquad\qquad \mathsf{M}_2(\tau) \quad = \quad \mathrm{extr}(\tau) \gg \mathrm{proj}(\tau).$$

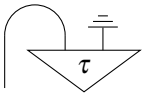
Proof We shall do the first equation and leave the others to the interested reader.

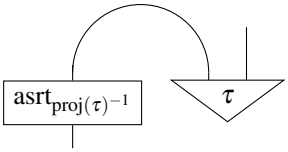
$$\begin{aligned}
\big(\mathrm{proj}(\mathrm{pair}(\sigma,c))\big)_{ij} &= \textstyle\sum_k \langle jk|\mathrm{pair}(\sigma,c)|ik\rangle \\
&\overset{(15)}{=} \textstyle\sum_k \overline{\big(\sqrt{\sigma}c_{kk}\sqrt{\sigma}\big)_{ji}} \\
&= \overline{\big(\sqrt{\sigma}\big(\textstyle\sum_k c_{kk}\big)\sqrt{\sigma}\big)_{ji}} \\
&= \overline{\sigma_{ji}} \qquad \text{since } c \text{ is unital, } \textit{i.e. } c \ll \mathbf{1} = \mathbf{1} \\
&= \big(\sigma^\dagger\big)_{ij} \\
&= \sigma_{ij}.
\end{aligned}$$

The latter equation holds since a state is positive and thus self-adjoint. □

As an aside, for readers who are comfortable with diagrammatic notation (see *e.g.* [25, 7]) one can write:

$\mathrm{pair}(\sigma,c) \;=$ 

$\qquad \mathrm{proj}(\tau) \;=$ 

$\qquad \mathrm{extr}(t) \;=$ 

6 A quantum Bayesian Inference Theorem

This section contains the main result of this paper, namely the quantum analogue of Theorem 2. It describes how conditioning of a joint state can also be performed via the extracted channel. The novelty in our quantum description is that we need both lower and upper conditioning to capture what is going on.

Theorem 3. *Let τ be a state of $\mathscr{H}\otimes\mathscr{K}$ and let p,q be predicates, on $\mathscr{H}$ and on $\mathscr{K}$ respectively. Then:*

$$\mathsf{M}_2\big(\tau|_{p\otimes\mathbf{1}}\big) \;=\; \mathrm{extr}(\tau) \gg \big(\mathrm{proj}(\tau)|^{p^T}\big) \qquad \textit{and} \qquad \mathsf{M}_1\big(\tau|_{\mathbf{1}\otimes q}\big) \;=\; \big(\mathrm{proj}(\tau)|^{\mathrm{extr}(\tau)\ll q}\big)^T. \quad \square$$

The proof is ommitted since it involves rather long and boring matrix calculations. Instead we include a random test: the quantum versions of pairing / projection / extraction and lower / upper conditioning have been implemented in EfProb. They can be used to test Theorem 3 as below, by generating an arbitrary state `t`, in this case of type $\mathbb{C}^3\otimes\mathbb{C}^5$, together with arbitrary (suitably typed) predicates. The EfProb notation for lower and upper conditioning is `/` and `^`.

```
>>> t = random_state([3,5])
>>> t1 = (t % [1,0]).transpose()
>>> e = extract(t)
>>> p = random_pred([3])
>>> q = random_pred([5])
>>> t / (p @ truth([5])) % [0,1] == e >> (t1 ^ p.transpose())
True
>>> t / (truth([3]) @ q) % [1,0] == (t1 ^ (e << q)).transpose()
True
```

The two equality tests == involve 5×5 and 3×3 matrices of complex numbers.

In the equations in Theorem 3 we perform lower conditioning on the joint state. One may ask if there are also 'dual' equations where upper conditioning on the joint state is re-described via state/predicate transformation. We have not found them.

Acknowledgements

Thanks to Kenta Cho and Alex Kissinger for helpful feedback and discussions.

References

[1] J.-M. Allen, J. Barrett, D. Horsman, C. Lee, and R. Spekkens. Quantum common causes and quantum causal models. *Phys. Rev. X*, 7(3):031021, 2017. doi:10.1103/PhysRevX.7.031021.

[2] D. Barber. *Bayesian Reasoning and Machine Learning*. Cambridge Univ. Press, 2012. doi:10.1017/CBO9780511804779. Publicly available via http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.HomePage.

[3] J. Bernardo and A. Smith. *Bayesian Theory*. John Wiley & Sons, 2000.

[4] K. Cho and B. Jacobs. Disintegration and Bayesian inversion, both abstractly and concretely. See arXiv:1709.00322, 2017.

[5] K. Cho and B. Jacobs. The EfProb library for probabilistic calculations. In F. Bonchi and B. König, editors, *Conference on Algebra and Coalgebra in Computer Science (CALCO 2017)*, volume 72 of *LIPIcs*. Schloss Dagstuhl, 2017. doi:10.4230/LIPIcs.CALCO.2017.25.

[6] K. Cho, B. Jacobs, A. Westerbaan, and B. Westerbaan. An introduction to effectus theory. See arXiv:1512.05813, 2015.

[7] B. Coecke and A. Kissinger. *Picturing Quantum Processes. A First Course in Quantum Theory and Diagrammatic Reasoning*. Cambridge Univ. Press, 2016. doi:10.1017/9781316219317.

[8] B. Coecke and R. Spekkens. Picturing classical and quantum Bayesian inference. *Synthese*, 186(3):651–696, 2012. doi:10.1007/s11229-011-9917-5

[9] B. Fong. Causal theories: A categorical perspective on Bayesian networks. Master's thesis, Univ. of Oxford, 2012. See arXiv:1301.6201.

[10] S. Gudder and R. Greechie. Sequential products on effect algebras. *Reports on Math. Physics*, 49(1):87–111, 2002. doi:10.1016/S0034-4877(02)80007-6.

[11] S. Hayashi. Adjunction of semifunctors: categorical structures in nonextensional lambda calculus. *Theor. Comp. Sci.*, 41:95–104, 1985. doi:10.1016/0304-3975(85)90062-3.

[12] R. Hoofman and I. Moerdijk. A remark on the theory of semi-functors. *Math. Struct. in Comp. Sci.*, 5(1):1–8, 1995. doi:10.1017/S096012950000061X.

[13] B. Jacobs. New directions in categorical logic, for classical, probabilistic and quantum logic. *Logical Methods in Comp. Sci.*, 11(3), 2015. doi:10.2168/LMCS-11(3:24)2015.

[14] B. Jacobs. From probability monads to commutative effectuses. *Journ. of Logical and Algebraic Methods in Programming*, 94:200–237, 2017. doi:10.1016/j.jlamp.2016.11.006

[15] B. Jacobs. Quantum effect logic in cognition. *Journ. Math. Psychology*, 81:1–10, 2017. doi:10.1016/j.jmp.2017.08.004.

[16] B. Jacobs and F. Zanasi. A predicate/state transformer semantics for Bayesian learning. In L. Birkedal, editor, *Math. Found. of Programming Semantics*, number 325 in Elect. Notes in Theor. Comp. Sci., pages 185–200. Elsevier, Amsterdam, 2016. doi:10.1016/j.entcs.2016.09.038.

[17] B. Jacobs and F. Zanasi. A formal semantics of influence in Bayesian reasoning. In K. Larsen, H. Bodlaender, and J.-F. Raskin, editors, *Math. Found. of Computer Science*, volume 83 of *LIPIcs*, pages 21:1–21:14. Schloss Dagstuhl, 2017. doi:10.4230/LIPIcs.MFCS.2017.21.

[18] B. Jacobs and F. Zanasi. The logical essentials of Bayesian reasoning. See arXiv:1804.01193, 2018.

[19] D. Koller and N. Friedman. *Probabilistic Graphical Models. Principles and Techniques*. MIT Press, Cambridge, MA, 2009.

[20] M. Leifer and R. Spekkens. Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference. *Phys. Rev. A*, 88(5):052130, 2013. doi:10.1103/PhysRevA.88.052130.

[21] M. Leifer and R. Spekkens. A Bayesian approach to compatibility, improvement, and pooling of quantum states. *Journ. of Physics A: Mathematical and Theoretical*, 47(27):275301, 2014. doi:10.1088/1751-8113/47/27/275301.

[22] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge Univ. Press, 2000. doi:10.1017/CBO9780511976667.

[23] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Graduate Texts in Mathematics 118. Morgan Kaufmann, 1988.

[24] J. Pienaar and Č. Brukner. A graph-separation theorem for quantum causal models. *New Journ. of Physics*, 17:073020, 2015. doi:10.1088/1367-2630/17/7/073020.

[25] P. Selinger. A survey of graphical languages for monoidal categories. In B. Coecke, editor, *New Structures in Physics*, number 813 in Lect. Notes Physics, pages 289–355. Springer, Berlin, 2011. doi:10.1007/978-3-642-12821-9_4