Picturing classical and quantum Bayesian inference

Bob Coecke*
Oxford University Computing Laboratory

Robert W. Spekkens†
Perimeter Institute

Abstract

We introduce a graphical framework for Bayesian inference that is sufficiently general to accommodate not just the standard case but also recent proposals for a theory of quantum Bayesian inference wherein one considers mixed quantum states rather than probability distributions as representative of degrees of belief. The diagrammatic framework is stated in the graphical language of symmetric monoidal categories and of compact structures and Frobenius structures therein, in which Bayesian inversion boils down to transposition with respect to an appropriate compact structure. In the case of quantum-like calculi, the latter will be non-commutative. We identify a graphical property that characterizes classical Bayesian inference. The abstract classical Bayesian graphical calculi also allow to model relations among classical entropies, and reason about these. We generalize conditional independence to this very general setting.

1 Introduction

In this paper we introduce a graphical calculus and corresponding axiomatics in terms of monoidal categories for a very general notion of Bayesian inference. It enables one to reason at a highly abstract level, about theories more general than 'classical' Bayesian inference, including earlier proposals for quantum Bayesian inference by Leifer [9] and by Leifer and Poulin [10]. The graphical language exploits the two-dimensional diagrammatic representation to distinguish givens and conclusions. Bayesian inversion is diagrammatic transposition in terms of the *compact structures* [7]. *Frobenius structures* [3] will be our vehicle for expressing notions such as *conditionalization* and relations of *conditional independence*. 'Classical' Bayesian inference is characterized in terms of a condition of commutativity for the Frobenius structure and therefore this structure is key to expressing *Bayesian updating* in the specific case of classical Bayesian inference.

An abstract representation of Bayesian inference allows one to identify which aspects of the standard probability calculus are merely conventional. For instance, in the context of R. T. Cox's derivation of the rules of classical Bayesian inference [6], the standard assumption that one's degree of belief about a proposition a ought to be represented by a number $p(a)$ between 0 and 1 and that one *multiplies* a conditional probability with a

*coecke@comlab.ox.ac.uk
†rspekkens@perimeterinstitute.ca

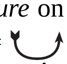
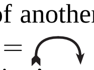
marginal to get the joint probability, i.e. $p(a,b) = p(a|b)p(b)$ is seen to be a consequence of a choice of convention. One could equally well represent this degree of belief by any bijective function of $p(a$ such as $s(a) = -\log p(a)$, in which case $s(a,b) = s(a|b) + s(b)$ and one replaces the standard form of Bayes' rule, $p(a|b) = p(b|a)p(a)/p(b)$, with its "entropic" form $s(a|b) = s(b|a) + s(a) - s(b)$. The abstract approach taken in this work finds a similar result and thereby contributes to the project of extracting the elements of Bayesian inference that are independent of convention.

Our graphical representation of Bayesian inference is also likely to have a close connection with the theory of Bayesian networks, and therefore may shed light on quantum analogues of these [10]. This has practical interest in the field of quantum information theory as quantum analogues of belief propagation algorithms are a natural avenue to quantum error correction schemes. As an example of this, the quantum analogue of Bayes' rule has the same form as the approximate reversal channel of Barnum and Knill [2].

2 Background: dagger Frobenius and compact structures

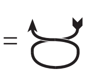
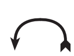
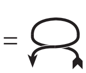
In this paper we work within the graphical language of symmetric monoidal categories (SMCs) [5, 14]. General morphisms (or operations) $f : A \to B$, which we interpret as 'processes', points (or elements) $e : \mathrm{I} \to A$, which we interpret as 'states', and composition and tensoring are respectively represented as:

(1)

A *compact structure* on an object A consists of another object A^* together with a pair of morphisms: $\eta_A =$  $: \mathrm{I} \to A^* \otimes A$ and $\epsilon_A =$  $: A \otimes A^* \to \mathrm{I}$, sometimes referred to as 'cups' and 'caps', which satisfy the 'yanking' equations:

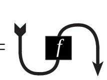

(2)

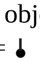
When we moreover have that $A^{**} = A$ then the direction of arrows clearly distinguishes between 'no $*$' and '$*$'. In this case, coherence requires us to set

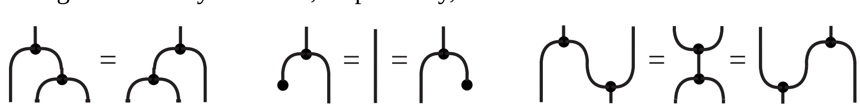
$\eta_{A^*} =$  $=$  $= \sigma_{A^*,A} \circ \eta_A$ $\quad$ $\epsilon_{A^*} =$  $=$  $= \epsilon_A \circ \sigma_{A^*,A}\,,$ (3)

where $\sigma_{A,B} : A \otimes B \to B \otimes A$ is the morphism that simply swaps the objects A and B.

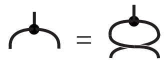
In any CC each morphism $f : A \to B$ has a *transpose*

$f^T := (1_{A^*} \otimes \epsilon_B) \circ (1_{A^*} \otimes f \otimes 1_{B^*}) \circ (\eta_A \otimes 1_{B^*}) =$  $: B^* \to A^*\,.$ (4)

A *dagger Frobenius structure* on an object A consists of an (internal) multiplication $m =$  $: A \otimes A \to A$ and a unit $u =$  $: \mathrm{I} \to A$ which satisfies the dagger Frobenius law. Diagrammatically these are, respectively,


(5)

The morphism $\delta := m^\dagger : A \to A \otimes A$ is called a *comultiplication* and $\epsilon := u^\dagger : A \to \mathrm{I}$ its *counit*. A dagger Frobenius structure is *commutative* when we have

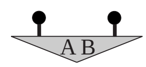
$m =$  $= m \circ \sigma_{A,A}\,.$ (6)

A dagger Frobenius structure admits an elegant diagrammatic calculus in terms of 'spider normal forms' [8] which we will use here extensively; e.g. see [4].

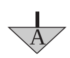
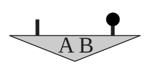
3 Bayesian graphical calculus

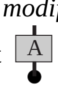
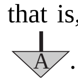
Consider a dSMC $\mathbf{C}$ in which each object comes with a dagger Frobenius structure.

BC1 For every object $A \in |\mathbf{C}|$, we assume the existence of a *normalized state*, that is, a point which when composed with the counit yields the morphism $1_\mathrm{I} : \mathrm{I} \to \mathrm{I}$ (the identity morphism on the trivial object), which we depict by the 'empty picture'  $=$ $: \mathrm{I} \to \mathrm{I}$. A normalized state for a composite object $A \otimes B \in |\mathbf{C}|$,

 $: \mathrm{I} \to A \otimes B$ such that  $=$ $: \mathrm{I} \to \mathrm{I}\,,$ (7)

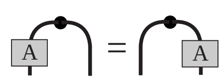
will be called a *joint state*. Note that the composition of a joint state on $A \otimes B$ with the counit on B is a state on A, which we call the *marginal state* on A,

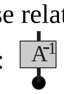
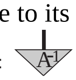
 $:=$  $: \mathrm{I} \to A\,.$ (8)

BC2 For every object $A \in |\mathbf{C}|$, we assume the existence of a *modifier*, that is, a self-transposed endomorphism  $: A \to A$ which is such that  $=$ .

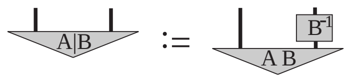
These modifiers are calculus-specific. We give concrete examples below of modifiers constructed in terms of marginal states and the Frobenius multiplication.

Proposition 3.1. *Since modifiers are self-transposed they can move along cups and caps:*

  . (9)

BC3 We assume that each state admits of a Frobenius inverse 'relative to its support' and each modifier admits an ordinary inverse relative to its support such that the latter is the modifier associated with the former:  $=$  .

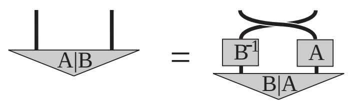
Definition 3.2. For every joint state on a pair of objects, the *conditional state* is



$: \mathrm{I} \to A \otimes B\,.$ (10)

We call a graphical calculus with ingredients **BC1**, **BC2**, **BC3** a *Bayesian graphical calculus*. This definition is motivated by the fact that with notions of joint states, marginal states, conditional states, modifiers and inverses, we have the minimal amount of structure required to describe basic concepts of Bayesian inference.

For example, Bayes' rule depicts as:

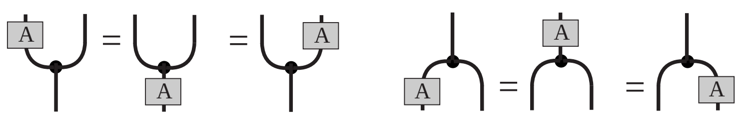


. (11)

Many important concepts can now be defined at this high level of generality, most notably, conditional independence, and many results can be derived graphically, e.g. pooling.

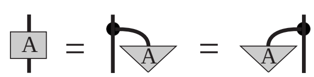
Definition 3.3. A Bayesian graphical calculus is called *classical* if it satisfies the following equivalent conditions:

(a) modifiers can move through the Frobenius structure:



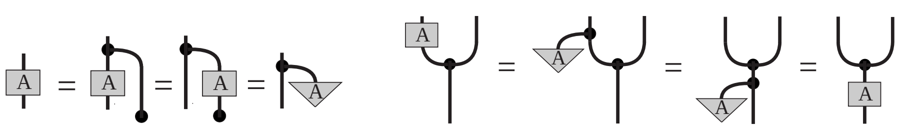
. (12)

(b) modifiers are of the form:



, (13)

The two conditions are related as follows:

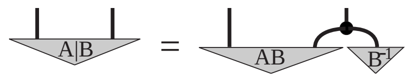


.

So in classical Bayesian graphical calculi, in addition to moving along cups and caps (cf. Proposition 3.1), modifiers can move through the Frobenius structure, and hence, by the spider theorem, in a classical Bayesian graphical calculus modifiers can move through arbitrary spiders. It is useful to consider some of the features of such a calculus.

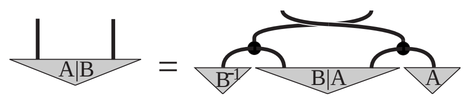
Proposition 3.4. *In a classical Bayesian graphical calculus, the Frobenius multiplication always acts commutatively on states and composition of modifiers is commutative.*

For a classical Bayesian calculus, conditional states have the form:



, (14)

and Bayes' theorem, Eq. (11), has the form



. (15)

By virtue of the multiplicative commutativity, the order in which the states are 'Frobenius-multiplied' doesn't matter (unlike the quantum generalization, as we will see).

This is an abstract characterization of classical Bayesian inference. We now present a couple of concrete realizations of this calculus. We shall thereby see how the abstract characterization avoids the conventional elements of the concrete realizations.

Example 3.5. Standard probability theory. Standard probability theory constitutes a special case of a classical Bayesian calculus. The objects are natural numbers and the morphisms from n to m are the $m \times n$ positive-valued matrices (consequently the points are column vectors and their daggers are row vectors). Composition is matrix product, and the tensor product is the matrix tensor product. The Bayes' rule for classical Bayesian graphical calculi as in Eq. (15), takes the form $p(A|B) = \frac{p(B|A)p(A)}{p(B)}$, where this is understood to be an equality that holds component by component.

Example 3.6. The negative logarithm of probability representation. Here everything is defined as it was before except that the underlying notions of scalar addition and multiplication are modified. The new operations, denoted by $\boxplus$ and $\boxdot$ respectively, can be defined for an arbitrary pair s, t of scalars as follows. For any function f that is bijective and hence invertible on the positive reals, they are

$$s \boxplus t = f(f^{-1}(s) + f^{-1}(t)), \qquad s \boxdot t = f(f^{-1}(s)f^{-1}(t)). \tag{16}$$

Consider the case where the monotonic function $f(s) = -\ln s$ and $f^{-1}(s) = e^{-s}$ so that

$$s \boxplus t = -\ln(e^{-s} + e^{-t}), \qquad s \boxdot t = s + t. \tag{17}$$

In this new calculus, an impossible value of k (one for which $p_k = 0$) is represented by $s_k = \infty$, while a certain value (one for which $p_k = 1$) is represented by $s_k = 0$. Now $s(A|B) = s(A, B) - s(B)$ and the Bayes' rule takes the form

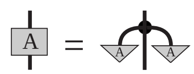
$$s(A|B) = s(B|A) + s(A) - s(B). \tag{18}$$

One has a choice in representing degrees of belief. It can be done with probabilities, but it can also be accomplished with negative logarithms of probabilities, or indeed any monotonic function of probabilities. It is a matter of convention only which is chosen. An argument to this effect was already made by R. T. Cox [6].

Example 3.7. Entropy. Taking the usual inner product of $s(A) := (s_1, s_2, \ldots, s_n)$ of negative logarithms of probabilities with $p(A) := (p_1, p_2, \ldots, p_n)$ of probabilities, one obtains the *Shannon entropy* of $p(A)$, $S(A) := \sum_k p_k s_k = -\sum_k p_k \ln p_k$, and similarly the joint entropy $S(A,B) := \sum_{i,j} p_{i,j} s_{i,j} = -\sum_{i,j} p_{i,j} \ln p_{i,j}$ and the conditional entropy $S(A|B) := \sum_{i,j} p_{i,j} s_{i|j} = -\sum_{i,j} p_{i,j} \ln p_{i|j}$. Noting the the marginal entropy can also be obtained by averaging over the joint distribution, $\sum_{i,j} p_{i,j} s_i = \sum_i p_i s_i = S(A)$, it follows that any expression that holds among joints, marginals and conditionals for negative logarithms of distributions (i.e. among $s_{i,j}$, s_i, $s_{i|j}$ etcetera) also holds among the joint, marginal and conditional *entropies*. For instance, Bayes' rule in terms of negative logarithms of probabilities, Eq. (18), implies the analogous relation among entropies

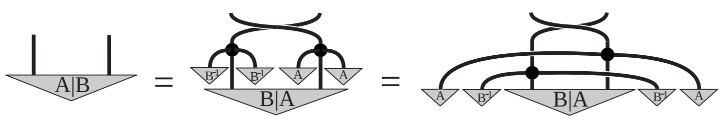
$$S(A|B) = S(B|A) + S(A) - S(B). \tag{19}$$

Definition 3.8. A Bayesian graphical calculus is a $Q_{1/2}$*-calculus* when modifiers are:



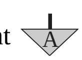
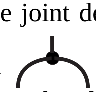
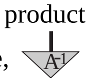
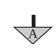
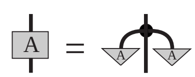
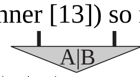
(20)

For $\mathrm{Q}_{1/2}$-calculi the Bayesian update law Eq. (11) becomes:



(21)

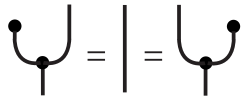
In the final expression of Eq. (21), the order of the two small triangles on the left could be reversed because they are not connected to each other by a spider. The same is true of the two small triangles on the right.

Example 3.9. The conditional density operator calculus. We take the point  to be a density operator $\rho(A) : A \to A$ and the point  to be the joint density operator $\rho(A,B) : A \otimes B \to A \otimes B$. We take the Frobenius multiplication  to be the (non-commutative) operator product $- \circ -$ of density operators, and hence the identity operator 1_A is its unit . Hence,  is the inverse density operator $\rho(A)^{-1}$,  is the square-root density operator $\sqrt{\rho(A)}$, and the modifier  is the completely positive map $\sqrt{\rho(A)} \circ - \circ \sqrt{\rho(A)}$. The trace is  (which is indeed the adjoint to the unit when taken in a suitable manner [13]) so marginals arise by tracing out a system on a joint density operator. The point  is Leifer's conditional density operator [9, 10], that is, a positive operator $\rho(A|B) : A \otimes B \to A \otimes B$ such that $\mathrm{Tr}_A[\rho(A|B)] = 1_B$.

Remark 3.10 (logical broadcasting). By a broadcasting operation we mean any operation $\delta : A \to A \otimes A$ acting on a space of density operators and satisfying

$$(tr_A \otimes 1_A) \circ \delta = 1_A = (1_A \otimes tr_A) \circ \delta\,. \tag{22}$$

A Frobenius comultiplication on density operators for which the operator trace $tr_A : A \to \mathrm{I}$ is its counit satisfies Eq. (22) by counitality:

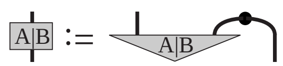


(23)

By the no-broadcasting theorem [1] it then also follows that such a Frobenius comultiplication is necessarily non-physical, i.e. it cannot be a completely positive map.

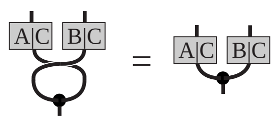
4 Inferential presentation of Bayesian graphical calculus

Above, we represented both joint and conditional states by the same triangles, only distinguishing them in terms of their labeling. We will now rely on the compact structure induced by the Frobenius structure to clearly distinguish between givens (objects on the right of the conditional bar "|" in our notation) and conclusions (objects on the left of the conditional) by representing the first as inputs (appearing at the bottom of the diagram) and the latter as outputs (appearing at the top). We define a *conditional process*:



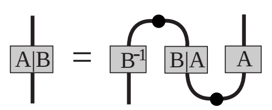
(24)

Proposition 4.1. *In a classical Bayesian graphical calculus:*



(25)

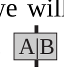
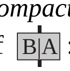
We shall refer to the diagrammatic representation of an expression wherein every conditional state is replaced by its isomorphic process as the *inferential presentation* because by reading the diagram from bottom to top one follows a chain of inferences. Bayes' rule for a general Bayesian calculus, described in Eq. (11), becomes:

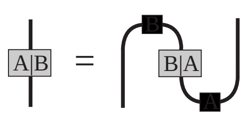


(26)

This form can be simplified further. One easily verifies that the morphisms

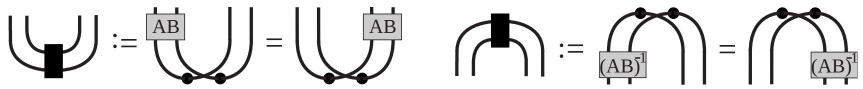
(27)

define another compact structure on A, which we will refer to as the *modified compact structure*. Bayes' rule is simply the statement that  is the *modified transpose* of  :



(28)

A modified compact structure for a pair of objects is



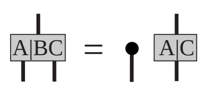
. (29)

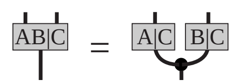
5 Conditional independence

In classical probability theory, a set of random variables X and another set Y are said to be *conditionally independent* given a third set Z if, equivalently:

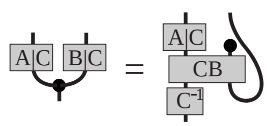
$$(a)\ p(X|Y,Z) = p(X|Z) \quad (b)\ p(Y|X,Z) = p(Y|Z) \quad (c)\ p(X,Y|Z) = p(X|Z)p(Y|Z)$$

In the general Bayesian graphical calculus, there are analogues of each of these conditions, but they are no longer equivalent. We therefore distinguish two pairs of notions of conditional independence. The first pair are the analogues of Eqs. (a) and (b) respectively, while the second pair constitute analogues of Eq. (c):

CI1 

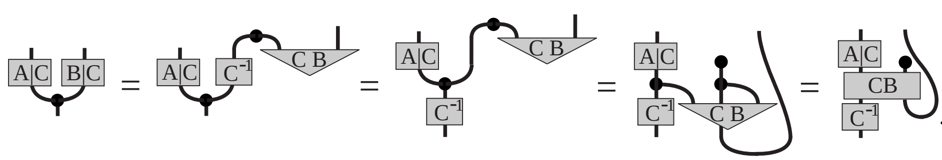
CI2 

Proposition 5.1. *In a Bayesian graphical calculus, if any two of* **CI1**, **CI2** *and* **F** *hold then the third one also holds, where* **F** *stands for:*

F 

Proposition 5.2. *In a classical Bayesian graphical calculus* **CI1** *and* **CI2** *are equivalent.*

Proof: The equality F always holds in a classical Bayesian graphical calculus:

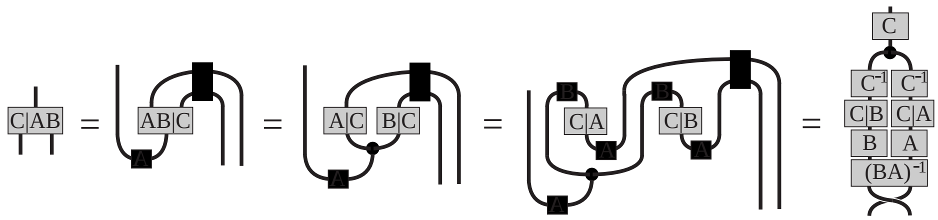


and hence by Proposition 5.1 **CI1** and **CI2** are equivalent. □

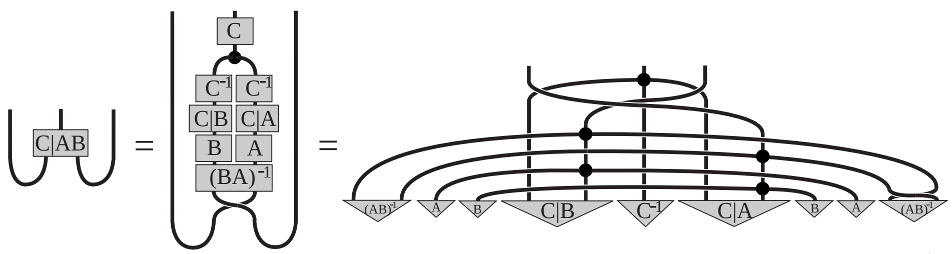
What is more difficult is to recover a quantum notion of conditional independence. An open question is whether specifying that the form of the modifiers is as given in Eq. (20) is sufficient to prove everything that can be proven within the conditional density operator calculus. In particular, it is not clear how to derive that **CI1** and **CI2** are equivalent.

A simple example of what one can derive from the notion of conditional independence, we consider the problem of pooling. Here, one seeks to assign a conditional state to C given A, B and the question is whether this state can be expressed in terms of a conditional state for C given A and a conditional state for C given B.

Proposition 5.3. *If A and B are conditionally independent relative to C as in* **CI2***, then*



For $Q_{1/2}$-calculi, when expressing this expression in terms of conditional states rather than in the inferential form we obtain:



(30)

For density operators, it is equivalent to

$$\rho(C|AB) = \sqrt{\rho(A,B)}^{-1}\sqrt{\rho(A)}\sqrt{\rho(B)}\rho(C|B)\rho(C)^{-1}\rho(C|A)\sqrt{\rho(B)}\sqrt{\rho(A)}\sqrt{\rho(A,B)}^{-1}$$

and for classical probability distributions, we obtain

$$P(C|AB) = \frac{P(A)\,P(B)}{P(A,B)} \cdot \frac{P(C|A)\,P(C|B)}{P(C)}\,. \tag{31}$$

This result is known as the *pooling formula* because if A and B are conditionally independent given C, the posterior $P(C|AB)$ can be reconstructed from the posteriors $P(C|A)$ and $P(C|B)$ and the prior $P(C)$ (the dependence on A and B is inferred from normalization). As such, it is sufficient to "pool" the information contained in the two posteriors.

6 Acknowledgement

The authors thank Matt Leifer for discussions and for feedback on a draft of this article, in particular, the suggestion to explore the graphoid axioms. RWS acknowledges support from the Perimeter Institute, which is funded by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Research and Innovation. BC acknowledges support from EPSRC ARF EP/D072786/1, ONR Grant N00014-09-1-0248, EU FP6 STREP QICS and Perimeter Institute who hosted him as a Long Term Visiting Scientist.

References

[1] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher (1996) *Non-commuting mixed states cannot be broadcast*. Physical Review Letters **76**, 2818–2821. arXiv:quant-ph/9511010

[2] H. Barnum and E. Knill (2002) *Reversing quantum dynamics with near-optimal quantum and classical fidelity*. Journal of Mathematical Physics **43**, 2097–2106.

[3] A. Carboni and R. F. C. Walters (1987) *Cartesian bicategories I*. Journal of Pure and Applied Algebra **49**, 11–32.

[4] B. Coecke (2010) *Quantum picturalism*. Contemporary Physics **51**, 59–83. arXiv:0908.1787

[5] B. Coecke and E. O. Paquette (2009) *Categories for the practising physicist*. In: New Structures for Physics, B. Coecke, Ed. Springer lecture Notes in Physics. To appear. arXiv:0905.3010

[6] R. T. Cox (1946) *Probability, frequency, and reasonable expectation*. American Journal of Physics **14**, 1–13.

[7] G. M. Kelly and M. L. Laplaza (1980) *Coherence for compact closed categories*. Journal of Pure and Applied Algebra **19**, 193–213.

[8] S. Lack (2004) *Composing PROPs*. Theory and Applications of Categories **13**, 147–163.

[9] M. S. Leifer (2006) *Quantum dynamics as an analog of conditional probability*. Physical Review A **74**, 042310. arXiv:quant-ph/0606022

[10] M. S. Leifer and D. Poulin (2008) *Quantum graphical models and belief propagation*. Annals of Physics **323**, 1899–1946. arXiv:0708.1337

[11] M. S. Leifer and R. W. Spekkens (2008) *Quantum analogues of Bayes' theorem, sufficient statistics and the pooling problem*, in preparation (2009).

[12] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Francisco, 1988.

[13] P. Selinger (2007) *Dagger compact categories and completely positive maps*. Electronic Notes in Theoretical Computer Science **170**, 139–163.

[14] P. Selinger (2010) *A survey of graphical languages for monoidal categories*. In: New Structures for Physics, B. Coecke (ed), pages 275–337. Lecture Notes in Physics, Springer-Verlag. arXiv:0908.3347