

# Copy-composition for probabilistic graphical models\*

Toby St Clere Smithe

VERSES Research Lab

Topos Institute

act@tsmithe.net

In probabilistic modelling, joint distributions are often of more interest than their marginals, but the standard composition of stochastic channels is defined by marginalization. Last year at ACT, the notion of ‘copy-composition’ was introduced in order to circumvent this problem and express the chain rule of the relative entropy fibrationally, but while that goal was achieved, copy-composition lacked a satisfactory origin story. Here, we supply such a story for two standard probabilistic tools: directed and undirected graphical models. We explain that (directed) Bayesian networks may be understood as “stochastic terms” of product type, in which context copy-composition amounts to a pull-push operation. Likewise, we show that (undirected) factor graphs compose by copy-composition. In each case, our construction yields a double fibration of decorated (co)spans. Along the way, we introduce a useful bifibration of measure kernels, to provide semantics for the notion of stochastic term, which allows us to generalize probabilistic modelling from product to dependent types.

## 1 Introduction

A submission to *Applied Category Theory 2023* formalized the “chain rule” of the relative entropy (*a.k.a.* Kullback-Leibler divergence) as a section of a fibration of statistical games over Bayesian lenses [1]. Curiously, however, the base category of this fibration was not a category of stochastic channels (measure kernels), as one might expect, but instead a bicategory of *copy-composite* channels. The need for this stemmed from the Chapman-Kolmogorov rule for the composition of kernels, which involves marginalization; and marginalization does not commute with the logarithm defining the relative entropy. Copy-composition defers this marginalization (indeed, the Chapman-Kolmogorov rule is precisely the marginalization of the intermediate copy), so there is no problem of commutativity, and the chain rule thus obtains.

Although this yields an elegant fibrational structure, copy-composition is not strictly unital, owing to the extra copy produced when copy-composing with the identity. This means that the base of the fibration is necessarily a bicategory. Moreover, the construction of this bicategory by St Clere Smithe [1] was largely *ad hoc*, based on the similarity to a **CoPara** construction. This was despite the evident importance of copy-composition to the statistical problems of interest there: copy-composition constructs *joint distributions*, which are often of more interest in applications than the composite channels that would yield their marginals. This can be seen not only via the relative entropy, but also in the context of Bayesian networks, which describe the factorization of a joint distribution according to a graphical structure. Fong [2] studied Bayesian networks from a categorical perspective, making extensive use of comonoid (‘copy-discard’ [3]) structure to construct the corresponding joint distributions; see for example the proof of his Theorem 4.5.

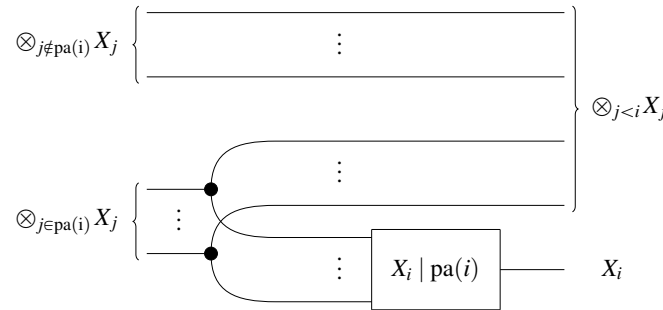
---

\*Some of this content appeared first on the author’s website, such as at <https://tsmithe.net/p/factor-graphs.html>; however, none has appeared elsewhere.

In the present submission, we show how copy-composition emerges naturally (but distinctly) from the compositional structure of directed and undirected probabilistic graphical models, thereby replacing the earlier arbitrariness with something structurally meaningful. In the directed case, the key observation is that a Bayesian network may be understood as a “stochastic term” of some displayed type (which is invariably a product), equipped with a distribution (the ‘prior’) over its context. Identifying types with bundles and terms with sections, this means understanding the factors of a Bayesian network as *stochastic sections*; copy-composition then emerges via a pull-push operation. In the undirected case, one observes that the ‘factors’ of a factor graph are costates<sup>1</sup> in a copy-discard category, and that copy-composition amounts to a kind of gluing. In both cases, copy-composition appears as the horizontal composition operation of a pseudo double category, thereby extending the earlier bicategorical story. These double categories<sup>2</sup> are in fact double fibrations, following the generalized decorated (co)span construction of Patterson [4]: in the directed case, one decorates spans of display maps with sections; in the undirected case, one decorates cospans with factors.

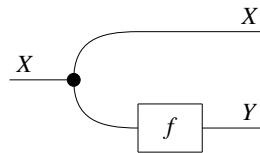
## 2 Directed models: pull-push stochastic sections

A *Bayesian network* is a distribution (or measure) on a product space that factorizes into a product of conditional distributions (or measure kernels) that matches the structure of a directed acyclic graph. Interpreting this idea in a category of stochastic channels (such as measure kernels between measurable spaces), Fong [2, Theorem 4.5] showed that any Bayesian network can be written as a state  $1 \rightarrow \otimes_j X_j$  that factorizes as the sequential composition of morphisms of the form



where  $\text{pa}(i)$  denotes the set of parents of the vertex  $i$ , and where the vertices may be given an ‘ancestral’ order  $j < i$  accordingly (such that  $j < i$  iff there is no directed path from  $i$  to  $j$ ).

Each factor  $X_i | \text{pa}(i)$  only appears in the composite diagram as precomposed by a copier applied to its parents, and tensored with the identities of its grandparents. Interpreting the analogous string diagram



in **Set**, we see that it corresponds to the graph of the function  $f : X \rightarrow Y$ , which is a section of the product projection  $X \times Y \rightarrow X$ ; in other words, a term of  $X \times Y$  over the context  $X$ . Moreover, ‘tensoring’ this

<sup>1</sup>A *state* in a monoidal category is a morphism out of the monoidal unit; a *costate* is a morphism into the monoidal unit.

<sup>2</sup>We take “double category” henceforth to mean *pseudo double category*, oriented such that composition is strict in the vertical category and weak in the horizontal.

diagram with extra objects  $W$  corresponds to pulling back the section along the projection  $W \times X \rightarrow X$ ; in other words, extending the context by  $W$ . Thus, whereas Fong [2] constructed the preceding diagram by hand, with the right categorical setting we will be able to interpret it as resulting from standard type-theoretic operations. Our first task will therefore be to establish a bifibration of measure kernels in which these operations may be interpreted. Then, we will construct a double category whose horizontal composition implements these operations (and thus copy-composition) via ‘pull-push’ in the bifibration.

## 2.1 A bifibration of measure kernels

The classic categorical semantics for dependent type theory is in fibrations: the base category is interpreted as a category of ‘contexts’; the fibre over an object interprets the types in the corresponding context; elements in a fibre interpret terms; and pullback (base change) interprets substitution. The classic example is the codomain fibration of a finitely complete category, in which bundles ‘display’ types, and their sections are corresponding terms. The category **Meas** of measurable spaces and measurable functions between them is a finitely complete category, and so this story may be told there. But measurable functions are still deterministic; they are not kernels. However, we may replace the functions in the fibres by kernels and still obtain a fibration, which is the purpose of this section. In fact, we will obtain a *bifibration*. We begin by making the notion of kernel more precise.

**Definition 2.1** (Measure kernel). Suppose  $X$  and  $Y$  are measurable spaces; we write  $\Sigma_X$  and  $\Sigma_Y$  for their  $\sigma$ -algebra structures. A (measure) kernel  $k : X \rightsquigarrow Y$  is a function  $k : X \times \Sigma_Y \rightarrow [0, \infty]$  that is measurable in its first argument and a measure in its second; *i.e.* such that  $k(-, B) : X \rightarrow [0, \infty]$  is measurable for all  $B \in \Sigma_Y$  and  $k(x, -) : \Sigma_Y \rightarrow [0, \infty]$  is a (positive) measure for all  $x \in X$ .

General kernels are too ill-behaved to yield a well-defined category, but for the purposes of §3, we will want more than just probability kernels (those that measure 1 on the whole codomain). Staton [5] showed that *s-finite* kernels (countable sums of finite kernels) are expressive enough for statistical modelling while also yielding a well-defined category, and it is these that we shall employ.

**Proposition 2.2** (Sums of kernels; [5, Proposition 1]). Suppose  $k_1, \dots, k_n, \dots$  is a countable sequence of kernels  $X \rightsquigarrow Y$ . Then there is a kernel  $\sum_i k_i : X \rightsquigarrow Y$  given by their sum,  $(\sum_i k_i)(x, B) := \sum_i k_i(x, B)$ .

**Definition 2.3** (Finite and s-finite kernels). Suppose  $k : X \rightsquigarrow Y$  is a kernel. Then:

1.  $k$  is *finite* if there is some  $r < \infty$  such that, for all  $x \in X$ ,  $k(x, Y) < r$ ;
2.  $k$  is *s-finite* if there is a countable sequence of finite kernels  $k_1, \dots$  such that  $k = \sum_i k_i$ .

Staton [5, Theorem 4] tells us that s-finite kernels satisfy Fubini’s theorem and thus yield a well-defined category. Composition of kernels  $X \rightsquigarrow Y$  and  $Y \rightsquigarrow Z$  is by Chapman-Kolmogorov; that is, by integrating over  $Y$ .

**Definition 2.4** (The category **sfKrn**). The objects of **sfKrn** are measurable spaces. Morphisms from  $X$  to  $Y$  are s-finite kernels  $X \rightsquigarrow Y$ . Given kernels  $X \xrightarrow{g} Y \xrightarrow{h} Z$ , their composite  $h \circ g : X \rightsquigarrow Z$  is defined by

$$h \circ g : (x, C) \mapsto \int_{y:Y} g(x, dy) h(y, C)$$

for  $x \in X$  and  $C \in \Sigma_Z$ . The identity kernel  $\text{id}_X : X \rightsquigarrow X$  is given by Dirac delta:  $\text{id}_X(x, -) := \delta_x(-)$ . Equivalently, let  $[x \in A]$  denote the indicator function

$$[x \in A] = \begin{cases} 1 & \text{if } x \in A; \\ 0 & \text{otherwise.} \end{cases}$$

Then  $\text{id}_X(x, A) = [x \in A]$ .

We will use **sfKrn** to define a bifibration  $\mathcal{K}$  over **Meas**, such that the objects of each fibre  $\mathcal{K}_B$  will be (deterministic, measurable) functions into  $B$ , and the morphisms  $(E, \pi) \rightsquigarrow (E', \pi')$  in  $\mathcal{K}_B$  will be kernels  $E \rightsquigarrow E'$  “fibrewise over  $B$ ”. To form this idea into a precise definition, we need to be able to compose (or *push forward*) a kernel with (or along) a deterministic function.

**Proposition 2.5** (Functions as kernels). Suppose  $f : Y \rightarrow Z$  is a measurable function. Then there is a kernel  $\delta_f : Y \rightsquigarrow Z$  defined by  $\delta_f(y, C) = [f(y) \in C]$ . This yields a functor  $\delta_{(-)} : \mathbf{Meas} \rightarrow \mathbf{sfKrn}$ .

**Definition 2.6** (Pushforward). Suppose  $k : X \rightsquigarrow Y$  is a kernel and  $f : Y \rightarrow Z$  is a measurable function. Then the *pushforward* of  $k$  along  $f$ , sometimes denoted  $f_*k$ , is the kernel  $X \rightsquigarrow Z$  defined by the composite  $\delta_f \circ k$  in **sfKrn**. Note that this composite satisfies  $(\delta_f \circ k)(x, C) = k(x, f^{-1}(C))$  where  $f^{-1}(C)$  is the preimage of  $C$  through  $f$ .

**Definition 2.7** (The fibres of  $\mathcal{K}$ ). Let  $B$  be a measurable space; we define a corresponding category  $\mathcal{K}_B$ . Its objects are pairs  $(E, p)$  of a measurable space  $E$  and a measurable function  $p : E \rightarrow B$ . A morphism  $k : (E, p) \rightarrow (E', p')$  is a kernel  $k : E \rightsquigarrow E'$  such that  $\delta_{p'} \circ k = \delta_p$ ; we call these morphisms *kernels over  $B$* . Composition of morphisms is given by the composition of the corresponding kernels; identities are given by identity kernels. (It is easy to see that this yields a well-defined category.)

To see that the morphisms of  $\mathcal{K}_B$  really are kernels fibrewise over  $B$ , let us introduce some notation.

**Notation 2.8** (Fibre notation). Suppose  $p : E \rightarrow B$  is a function and  $b : J \rightarrow B$  is a generalized element of  $B$  (another function). The fibre of  $p$  over  $b$  is given by the pullback of  $p$  along  $b$ :

$$\begin{array}{ccc} p[b] & \xrightarrow{\pi_E} & E \\ b^*p \downarrow & \lrcorner & \downarrow p \\ J & \xrightarrow{b} & B \end{array}$$

We denote the fibre by  $p[b] \xrightarrow{b^*p} J$ .

**Proposition 2.9** (Morphisms in  $\mathcal{K}_B$  are fibrewise kernels). Suppose  $k : (E, p) \rightsquigarrow (E', p')$  is a kernel over  $B$ , and  $b : J \rightarrow B$  is a generalized element of  $B$ . Then  $k$  restricts to a kernel  $k[b] : p[b] \rightsquigarrow p'[b]$ .

*Proof.* Let  $\pi_E : p[b] \rightarrow E$  be the projection as in the preceding pullback square; and let  $\pi_{E'} : p'[b] \rightarrow E'$  be likewise for  $p'$ . Then we have a kernel  $k[b] : p[b] \times_{\Sigma_{p'[b]}} \rightarrow [0, \infty]$  by setting  $k[b](e_b, U_b) := k(\pi_E(e_b), \pi_{E'}(U_b))$ , where  $\pi_{E'}(U_b)$  denotes the image of  $U_b$  under  $\pi_{E'}$ .  $\square$

It will be very useful to extend this fibrewise-restriction to composite kernels; fortunately, it commutes with composition. For the proof, see §A.1.

**Proposition 2.10** (Fibrewise restriction commutes with composition).  $(k \circ h)[b] = k[b] \circ h[b]$ , whenever  $k \circ h$  exists.

It is moreover easy to check that fibrewise restriction preserves identity kernels, which, together with the preceding proposition, means that it is functorial. In particular, given  $b : J \rightarrow B$  in **Meas**, fibrewise restriction yields a reindexing functor  $\Delta_b : \mathcal{K}_B \rightarrow \mathcal{K}_J$ .

**Corollary 2.11** (Substitution). Suppose  $b : J \rightarrow B$  in **Meas**. Then we have a functor  $\Delta_b : \mathcal{K}_B \rightarrow \mathcal{K}_J$  as follows. On objects,  $\Delta_b$  acts to map  $E \xrightarrow{p} B$  to  $p[b] \xrightarrow{b^*p} J$ . On morphisms, it maps  $k$  to  $k[b]$ .

**Proposition 2.12** (The stochastic self-indexing, and resulting fibration). Together with the substitution functors  $\Delta$ , the mapping  $\mathcal{K} : B \mapsto \mathcal{K}_B$  defines a pseudo functor  $\mathbf{Meas}^{\text{op}} \rightarrow \mathbf{Cat}$ . Pseudo functoriality follows from the (pseudo) functoriality of pullback and the functoriality of composition. Applying the Grothendieck construction to this indexed category yields a fibration  $\int \mathcal{K} \rightarrow \mathbf{Meas}$ .

It remains to exhibit left adjoints to the substitution functors  $\Delta_b$ .

**Proposition 2.13** (Dependent sum functor). Suppose  $f : C \rightarrow B$  is a morphism in **Meas**. Then there is a functor  $\Sigma_f : \mathcal{K}_C \rightarrow \mathcal{K}_B$  defined on objects  $E \xrightarrow{p} C$  by post-composition,  $\Sigma_f(p) := E \xrightarrow{p} C \xrightarrow{f} B$  and on kernels as the identity.

Note that, fibrewise, we have  $\Sigma_f(p)[b] \cong p[f^*b]$ . Since  $f^*$  is pullback in **Meas** and  $\Delta$  is defined by pullback, this is the first sign of the adjointness of  $\Sigma$  and  $\Delta$ . See §A.1 for the proof of the proposition:

**Proposition 2.14** ( $\Sigma \dashv \Delta$ ). Suppose  $f : B \rightarrow C$  in **Meas**. Then  $\Sigma_f$  is left adjoint to  $\Delta_f$ .

Thus  $\mathcal{K}$  yields a bifibration. But this is not all we need: for copy-composition to be well-defined via pull-push, we need the functors  $\Sigma$  and  $\Delta$  to satisfy the Beck-Chevalley condition. This result follows because the self-indexing of **Meas** satisfies Beck-Chevalley; but an explicit proof may be found in §A.1.

**Proposition 2.15** ( $\mathcal{K}$  satisfies Beck-Chevalley). Suppose we have a pullback square in **Meas** of the form

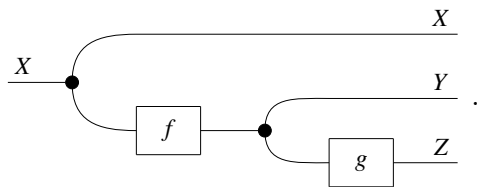
$$\begin{array}{ccc}
 & P & \\
 \pi \swarrow & \downarrow & \searrow \rho \\
 E & & F \\
 p \searrow & & \swarrow q \\
 & B &
 \end{array}$$

Then there is a natural isomorphism  $\Sigma_p \Delta_\pi \cong \Delta_q \Sigma_\rho$ .

In fact, the bifibration  $\int \mathcal{K}$  has more structure than we have space to explore in the present study: each fibre has a tensor product that makes it into a monoidal fibration; each fibre additionally has biproducts (and, with an inconsequential modification, a zero object); and thus the subcategory of probability kernels in each fibre is a *commutative effectus* [6]. Moreover, these structures interact nicely with the base change functors. We leave the exposition of all this to future work.

## 2.2 Decorating spans with sections

Our task now is to define a categorical structure in which morphisms include graphs (in the functional sense) and compose by copy-composition. That is, given the ‘graphs’ of  $f : X \rightarrow Y$  and  $g : Y \rightarrow Z$  in some underlying monoidal category, we wish to be able to compose them to yield



This composite is a section of the composite projection  $X \times Y \times Z \rightarrow X \times Y \rightarrow X$ . In turn, we may observe that this projection is the left leg of the composite span

$$\begin{array}{ccccc}
 & & X \times Y \times Z & & \\
 & \swarrow & \downarrow & \searrow & \\
 \pi_X^{XY} \swarrow & X \times Y & & Y \times Z & \searrow \pi_Z^{YZ} \\
 & \swarrow \pi_X^{XY} & & \swarrow \pi_Y^{YZ} & \\
 X & & Y & & Z
 \end{array}$$

and that the graphs of  $f$  and  $g$  are sections of the left legs of the constituent spans, respectively. Thus, if we can interpret  $\text{graph}(f)$  and  $\text{graph}(g)$  bifibrationally, we could write their composite (depicted above) as  $\Sigma_{\pi_X^{XY}} \Delta_{\pi_Y^{XY}} (\text{graph}(g)) \circ \text{graph}(f)$ . The operation  $\Sigma_{\pi_X^{XY}} \Delta_{\pi_Y^{XY}}$  is known as a “pull-push” operation [7], because one first pulls back along  $\pi_Y^{XY}$  and then pushes forward along  $\pi_X^{XY}$  (but this is a different pushing-forward to that of Definition 2.6).

When  $f$  and  $g$  are morphisms in a finitely complete category, we can immediately make sense of this pull-push composition in the codomain fibration. Likewise, when they are s-finite kernels, we can interpret this story in  $\mathcal{K}$ , and this is the situation that allows us to construct Bayesian networks and recover the chain rule of the relative entropy. Nonetheless, all we require is a bifibration satisfying Beck-Chevalley (and a section of it). The idea is then to decorate spans in the base of the bifibration with sections of the left leg, and then to compose them by pullback on the spans and pull-push on the decorations. Thus we have a “decorated span” construction, in the sense of Fong [8] and Patterson [4].

Patterson [4] observed that adding compositional decorations to categorical constructions often signals a Grothendieck fibration. Since spans (and cospans) collect into double categories, the decorated (co)span construction can be reconceived as a double Grothendieck construction, giving an extra dimension of composition to that originally proposed by Fong [8], and newly allowing the decorations to depend on whole (co)spans (rather than merely their apices), a freedom we will need. Following Cruttwell et al. [9, Theorem 3.51], the double Grothendieck construction produces a double fibration from an indexed double category, *i.e.* a lax double pseudo functor whose codomain is the double 2-category  $\mathbf{Span}(\mathbf{Cat})$  of spans of categories; see Patterson [4, §3.1] for a summary.

**Definition 2.16.** A lax double pseudo functor  $F : \mathbb{C} \rightarrow \mathbb{D}$  consists of a pair of pseudo functors  $F_0 : \mathbb{C}_0 \rightarrow \mathbb{D}_0$  (from the vertical category of  $\mathbb{C}$  to that of  $\mathbb{D}$ ) and  $F_1 : \mathbb{C}_1 \rightarrow \mathbb{D}_1$  (from the horizontal category of  $\mathbb{C}$  to that of  $\mathbb{D}$ ), along with pseudo natural transformations  $\mu$  and  $\eta$  witnessing the compatibility of  $F_1$  and  $F_0$  with the double category structures of  $\mathbb{C}$  and  $\mathbb{D}$ , as in the diagrams

$$\begin{array}{ccc}
 \mathbb{C}_1 \times_{\mathbb{C}_0} \mathbb{C}_1 & \xrightarrow{\diamond_{\mathbb{C}}} & \mathbb{C}_1 \\
 F_1 \times F_1 \downarrow & \nearrow \mu & \downarrow F_1 \\
 \mathbb{D}_1 \times_{\mathbb{D}_0} \mathbb{D}_1 & \xrightarrow{\diamond_{\mathbb{D}}} & \mathbb{D}_1
 \end{array}
 \qquad
 \begin{array}{ccc}
 \mathbb{C}_0 & \xrightarrow{I_{\mathbb{C}}} & \mathbb{C}_1 \\
 F_0 \downarrow & \nearrow \eta & \downarrow F_1 \\
 \mathbb{D}_0 & \xrightarrow{I_{\mathbb{D}}} & \mathbb{D}_1
 \end{array}$$

such that four standard coherence axioms [9, Def. 3.12] are satisfied. ( $\diamond$  denotes the ‘external’ (2-cell) composition of a double category structure and  $I$  its unit, so that, in the context of indexed double categories,  $\mu$  implements the *indexed* external composition, and  $\eta$  implements *its* unit.)

An *indexed double category* is a lax double pseudo functor with codomain  $\mathbf{Span}(\mathbf{Cat})$ .

Suppose now that  $\mathcal{E} \xrightarrow{\pi} \mathcal{B}$  is a bifibration satisfying Beck-Chevalley, equipped with a section  $\iota : \mathcal{B} \rightarrow \mathcal{E}$ , where  $\mathcal{B}$  has all pullbacks; we will continue to write  $\mathcal{E}_B$  for the fibre over  $B : \mathcal{B}$ , and  $\Sigma \dashv \Delta$  for the base change functors, and assume that  $\Delta$  preserves  $\iota$ . Our model for  $\mathcal{E}$  will be either  $\int \mathcal{K}$  or the codomain fibration of a finitely complete category; and for  $\iota$  the mapping  $B \mapsto (\text{id}_B : B \rightarrow B)$ , so that morphisms out of  $\iota$  are understood as sections (and thus models for terms). We will construct an indexed double category  $\mathcal{S} : \mathbf{Span}(\mathcal{B})_{pb}^{\text{op}} \rightarrow \mathbf{Span}(\mathbf{Cat})$ , where  $\mathbf{Span}(\mathcal{B})_{pb}$  is the sub-double category of  $\mathbf{Span}(\mathcal{B})$  whose 2-cells are Cartesian morphisms of spans (a restriction necessary for morphisms of decorations to be well-defined), and the  $^{\text{op}}$  is taken vertically.

The vertical category of  $\mathbf{Span}(\mathcal{B})_{pb}$  is  $\mathcal{B}$ , and the vertical 2-category of  $\mathbf{Span}(\mathbf{Cat})$  is  $\mathbf{Cat}$ ; but the vertical decorations will be trivial:  $S_0$  is simply the constant functor on the terminal category  $\mathbf{1}$ .

There is a simplification in the horizontal direction, too. The horizontal category of  $\mathbf{Span}(\mathcal{B})_{pb}$  is  $\mathcal{B}_{pb}^{\{\bullet \leftarrow \bullet \rightarrow \bullet\}}$ , the category of diagrams in  $\mathcal{B}$  of shape  $\bullet \leftarrow \bullet \rightarrow \bullet$ , with morphisms restricted to those natural transformations that are Cartesian (*i.e.*, constituting pullback squares).  $S_1$  must map a span  $m = (A \xleftarrow{a} E \xrightarrow{b} B)$  in  $\mathcal{B}$  to a span  $\mathbf{1} \leftarrow S_1(m) \rightarrow \mathbf{1}$  of categories. Because  $\mathbf{1}$  is terminal in  $\mathbf{Cat}$ , the legs of the latter span are accordingly determined. There is a further simplification:  $S_1(m)$  will be a discrete category (a set). Indeed, we define  $S_1(m)$  to be the set of sections of the left leg of  $m$  in  $\mathcal{E}$ . That is,  $S_1(m) := \mathcal{E}_A(\iota A, \Sigma_a \iota E)$ .

Given a morphism  $f : m \rightarrow m'$  in  $\mathbf{Span}(\mathcal{B})_{pb}$  as in the diagram

$$\begin{array}{ccccc} A & \xleftarrow{a} & E & \xrightarrow{b} & B \\ f_l \downarrow & & \downarrow & & \downarrow f_r \\ A' & \xleftarrow{a'} & E' & \xrightarrow{b'} & B' \end{array},$$

$S_1(f)$  must be a function  $\mathcal{E}_{A'}(\iota A', \Sigma_{a'} \iota E') \rightarrow \mathcal{E}_A(\iota A, \Sigma_a \iota E)$ , which we take to be given by  $\Delta_{f_l}$ . This is well defined because  $\Delta_{f_l} \iota A' = \iota A$  *ex hypothesi*,  $\Delta_{f_l} \Sigma_{a'} \iota E' \cong \Sigma_a \Delta_{f_l} \iota E'$  by Beck-Chevalley, and  $\Delta_{f_l} \iota E' = \iota E$  *ex hypothesi*. Moreover, since  $\Delta_{f_l}$  is functorial, so is  $S_1(f)$ .

The copy-composition of sections is implemented by the transformation  $\mu$ , whose component at a pair of composable spans  $m = (A \xleftarrow{a} E \xrightarrow{b} B)$  and  $n = (B \xleftarrow{b'} F \xrightarrow{c} C)$  must be a function  $\mu_{m,n} : S_1(m) \times S_1(n) \rightarrow S_1(n \diamond m)$ , which we take to be defined by pull-push. That is,  $\mu_{m,n}(\sigma, \tau) := \Sigma_a \Delta_b(\tau) \circ \sigma$ . Associativity of  $\mu$  follows by Beck-Chevalley; the explicit proof is in Appendix A.2. The unit  $\eta$  is defined over each identity span  $E = E = E$  by the function  $\eta_E : \mathbf{1} \rightarrow \mathcal{E}_E(\iota E, \iota E)$  determined by the identity on  $\iota E$ . Note that this makes pull-push copy-composition strictly unital, unlike in St Clere Smithe [1].

Applying the double Grothendieck construction [9, Theorem 3.51] to  $S$  yields a double category  $\mathbb{S}$ , doubly fibred over  $\mathbf{Span}(\mathcal{B})_{pb}$ , which has the following structure:

1. The objects of  $\mathbb{S}$  are the objects of  $\mathcal{B}$ .
2. A vertical 1-cell is a morphism of  $\mathcal{B}$ , and vertical composition is as in  $\mathcal{B}$ .
3. A horizontal 1-cell  $A \rightarrow B$  is a quadruple  $(E, a, b, \sigma)$  where  $(A \xleftarrow{a} E \xrightarrow{b} B)$  is a span with  $\sigma$  being a section of the left leg in  $\mathcal{E}_A$ ; *i.e.*,  $\sigma \in \mathcal{E}_A(\iota A, \Sigma_a \iota E)$ .
4. A 2 cell from  $(E, a, b, \sigma) : A \rightarrow B$  to  $(E', a', b', \sigma') : A' \rightarrow B'$  is a Cartesian morphism of spans  $(f_l, f, f_r) : (A \xleftarrow{a} E \xrightarrow{b} B) \Rightarrow (A' \xleftarrow{a'} E' \xrightarrow{b'} B')$  such that  $\sigma = \Delta_{f_l} \sigma'$ .
5. Horizontal composition is by pullback on the spans (and their morphisms) and by pull-push on the sections, as described above.

That is to say, we have the following theorem (given the details in Appendix A.2).

**Theorem 2.17** (Copy-composition of sections). Given a bifibration  $\mathcal{E} \rightarrow \mathcal{B}$  over a finitely complete base, satisfying Beck-Chevalley, and equipped with a section  $\iota : \mathcal{B} \rightarrow \mathcal{E}$  preserved by pullback, the foregoing data define a pseudo double category  $\mathbb{S}$  doubly fibred over  $\mathbf{Span}(\mathcal{B})_{pb}$ .

By instantiating  $\mathbb{S}$  in  $\int \mathcal{K}$ , we can formally recover the factors of a Bayesian network described by Fong [2] and introduced at the beginning of this section. There is a lax embedding of  $\mathbf{sfKrn}$  horizontally into  $\mathbb{S}$  (so instantiated), which maps a kernel to its graph. Let  $\sigma$  denote the image of the kernel  $X_i \mid \text{pa}(i)$  under this embedding, which has the effect of precomposing it with copiers. Then, to tensor with the

identity wires  $\otimes_{j \notin \text{pa}(i)} X_j$ , we can pull  $\sigma$  back along the projection  $\prod_{j < i} X_j \rightarrow \prod_{j \in \text{pa}(i)} X_j$ , as in the diagram

$$\begin{array}{ccccc}
 & & \xrightarrow{\sigma_{\leq i}} & & \\
 \prod_{j < i} X_j & \longleftarrow & \prod_{j \leq i} X_j & \xlongequal{\quad} & \prod_{j \leq i} X_j \\
 \downarrow & & \downarrow & & \downarrow \\
 \prod_{j \in \text{pa}(i)} X_j & \longleftarrow & \prod_{j \in \text{pa}(i)} X_j \times X_i & \longrightarrow & X_i
 \end{array}$$

The upper decorated span is the horizontal morphism corresponding to Fong’s factor. Finally, note that whereas traditional statistical modelling is restricted to sections of product types (as above), using  $\mathcal{K}$  and  $\mathbb{S}$  means we can build probabilistic models involving general dependent types.

### 3 Undirected models: open factor graphs

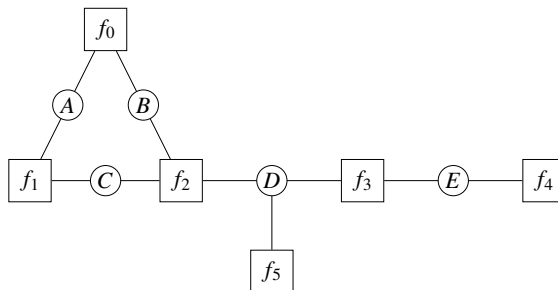


Figure 1: A factor graph, in the style of Wainwright and Jordan [10, §2.1.3].

The preceding section explained how copy-composition is central to Bayesian networks: *directed* probabilistic graphical models. In this section, we shall see that it is also central to the construction of undirected models known as *factor graphs* [10, 11]. A factor graph is a graphical representation of the factorization of a function, which is often but not necessarily interpreted as the density of a probability distribution. Thus Figure 1 represents a function  $f$  of the form

$$f(a, b, c, d, e) = f_0(a, b) f_1(a, c) f_2(b, c, d) f_3(d, e) f_4(e) f_5(d)$$

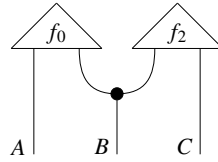
where the variables have the types  $a : A, b : B, c : C, d : D$ . Note that we could see  $f$  as composed of ‘factors’  $f_i$ ; it is this composition that we explore here.

Typically, the factors of a factor graph are functions into some object of scalars; for example, we might interpret  $f_2$  (above) to have the type  $B \times C \times D \rightarrow \mathbb{R}_+$ , for some spaces  $B, C, D$ . Should these spaces be  $\mathbb{R}_+$ -vector spaces, then we might suppose  $f_2$  to be a linear functional or *covector* on their product.  $\mathbb{R}_+$ -vector spaces form a category  $\mathbf{Vect}_{\mathbb{R}_+}$  which is equipped with a tensor product  $\otimes$ , whose unit is the 1-dimensional space of scalars  $\mathbb{R}_+$ , often denoted  $I$ . Covectors on  $X$  then correspond to *costates*, morphisms  $X \rightarrow I$ . This suggests that we may internalize the notion of factor in a monoidal category  $(\mathcal{C}, \otimes, I)$  simply as the costates in  $\mathcal{C}$ .

This first step towards an abstract treatment should be taken with care, as any semicartesian monoidal category, such as  $\mathbf{Set}$  or any Markov category (in the sense of Fritz [12]), will only have trivial costates, the monoidal unit there being by definition terminal. The composition of factors then proceeds by



copy-composition; notice for example the duplication of the variable  $b : B$  in the expression for  $f$  above, or in the following string-diagrammatic representation of its  $f_0, f_2$  factor:



This suggests that an appropriate categorical setting for factor graphs is that of *copy-discard categories*<sup>3</sup>, of which **sfKrn** (or any fibre of  $\mathcal{K}$ ) is an example.

Because factor graphs are a form of undirected graphical model, it is natural to expect them to fit into the framework of undirected wiring diagrams [13] [14, Chapter 7]. Undirected wiring diagrams are algebras for an operad of cospans [15], and one way to obtain such algebras is through the decorated cospans construction, originally due to Fong [8]. (But note that the decorated spans of the previous section are not undirected in the same way, because the decorations there are asymmetric.) The main result of this section is to show that ‘open’ factor graphs may be obtained and composed precisely through the double-categorical generalization of this construction due to Patterson [4].

**Remark 3.1.** Fong and Spivak [15] show that cospan-algebras are equivalent to hypergraph categories, in which each morphism has the form of a hypergraph. Consequently, this is also the case for factor graphs, meaning that the traditional name is something of a misnomer.

The remainder of this section is dedicated to constructing a double category indexed by cospans, *i.e.* a double pseudo functor  $F$ , and instantiating its double Grothendieck construction, the double category  $\mathbb{F}\mathbb{G}$ . The domain of  $F$  is the double category of cospans of finite sets, whose vertical category  $\mathbf{Cospan}(\mathbf{FinSet})_0$  is simply  $\mathbf{FinSet}$ , and whose horizontal category  $\mathbf{Cospan}(\mathbf{FinSet})_1$  is the category of diagrams in  $\mathbf{FinSet}$  of shape  $\bullet \rightarrow \bullet \leftarrow \bullet$ , *i.e.* the functor category  $\mathbf{FinSet}^{\{\bullet \rightarrow \bullet \leftarrow \bullet\}}$ . The codomain of  $F$  is  $\mathbf{Span}(\mathbf{Cat})$ , whose vertical 2-category is  $\mathbf{Cat}$ , and whose horizontal 2-category is  $\mathbf{Cat}^{\{\bullet \leftarrow \bullet \rightarrow \bullet\}}$ .

Henceforth, we will assume an ambient copy-discard category  $(\mathcal{C}, I, \otimes)$ . Because finite sets are not ordered and we plan to define factors over accordingly indexed monoidal products, we must additionally assume that the monoidal structure  $(I, \otimes)$  on  $\mathcal{C}$  is symmetric.

There will be fewer simplifications involved in the definition of  $F$  (versus  $S$ ), and so we split the construction into four subsections.

### 3.1 Vertical decoration: factors’ interfaces

The pseudo functor  $F_0 : \mathbf{FinSet} \rightarrow \mathbf{Cat}$  assigns categories of decorations to finite sets. For us, these decorations will be the exposed interfaces of factors, given by a discrete diagram in  $\mathcal{C}$ : a functor  $\text{disc}X \rightarrow \mathcal{C}$ , for some finite set  $X$ , where  $\text{disc}X$  is the discrete category on  $X$ . However, were we just to define  $F_0X$  to be the functor category  $\mathbf{Cat}(\text{disc}X, \mathcal{C})$ , we would run into problems later with the composition of 2-cells. This is because vertical morphisms — decorated by morphisms in the image of  $F_0$  — may be used to transform the interface types, but the composition of factors proceeds by copying. For this to be compatible with morphisms of interfaces, we must restrict the latter to be comonoid homomorphisms.

<sup>3</sup>A *copy-discard category* is a monoidal category in which every object is equipped with a comonoid structure (with respect to the ambient monoidal structure) [3]. A *Markov category* is a semicartesian copy-discard category [12]. Precisely because of the existence of non-trivial costates, copy-discard categories often end up being more useful in modelling practice than Markov categories, despite the utility of the latter for synthetic probability theory.

Therefore, on objects  $X : \mathbf{FinSet}$ , let  $F_0X$  be the functor category  $\mathbf{Cat}(\text{disc}X, \mathbf{Comon}(\mathcal{C}, I, \otimes))$  where  $\mathbf{Comon}(\mathcal{C}, I, \otimes)$  denotes the (wide) monoidal subcategory of  $\mathcal{C}$  whose morphisms are the comonoid homomorphisms. Note that, because the objects  $\chi$  of  $F_0X$  are functors on a discrete category, a morphism  $\varphi : \chi \rightarrow \chi'$  (hence a natural transformation) is simply given by an  $X$ -indexed family of comonoid homomorphisms  $\varphi_x : \chi(x) \rightarrow \chi'(x)$  in  $\mathcal{C}$ .

Given a morphism  $f : X \rightarrow Y$  in  $\mathbf{FinSet}$ , we let  $F_0$  return a functor  $F_0X \rightarrow F_0Y$  which we denote by  $f_*$  and define as follows. If  $\chi$  is an object of  $F_0X$  then we define  $f_*\chi$  by the mapping  $y \mapsto \otimes_{x:f^{-1}(y)} \chi(x)$  (for  $y : Y$ ). Then, if  $\varphi : \chi \rightarrow \chi'$  is a morphism of  $F_0X$ , we define each  $y$ -component  $(f_*\varphi)_y$  of  $f_*\varphi$  to be  $\otimes_{x:f^{-1}(y)} \varphi_x$ . Note that this definition yields components of the correct type

$$(f_*\varphi)_y : (f_*\chi)(y) = \otimes_{x:f^{-1}(y)} \chi(x) \xrightarrow{\otimes_{x:f^{-1}(y)} \varphi_x} \otimes_{x:f^{-1}(y)} \chi'(x) = (f_*\chi')(y).$$

Naturality is trivially satisfied, because  $\chi$  and  $\chi'$  are discrete functors, and the functoriality of the reindexing  $f_*$  follows from the functoriality of  $\otimes$ . Likewise, the pseudofunctoriality of  $F_0$  itself follows from the functoriality of inverse images and the functoriality and symmetry of  $\otimes$ , so that  $\otimes_{y:g^{-1}(z)} \otimes_{x:f^{-1}(y)} \cong \otimes_{x:(g \circ f)^{-1}(z)}$ .

### 3.2 Horizontal decoration: ‘open’ factors

The pseudo functor  $F_1 : \mathbf{FinSet}^{\{\bullet \rightarrow \bullet \leftarrow \bullet\}} \rightarrow \mathbf{Cat}^{\{\bullet \leftarrow \bullet \rightarrow \bullet\}}$  assigns spans of categories to cospans of finite sets, compatibly with  $F_0$ . We interpret a cospan  $A \xrightarrow{a} X \xleftarrow{b} B$  as follows. The apex set  $X$  will be decorated with two kinds of data: first, a category of interfaces suited to factors in  $\mathcal{C}$ ; and second, for each suitable interface, a category of factors on that interface (compatibly with morphisms of interfaces). The legs of the cospan  $a$  and  $b$  will be used to expose those parts of the interface which are open to composition, and in this way the objects  $A$  and  $B$  will be decorated by  $F_0A$  and  $F_0B$  respectively; this is one of the indexed-double-categorical coherence axioms (‘well-definition’).

Thus, given a cospan  $m = (A \xrightarrow{a} X \xleftarrow{b} B)$ , we let  $F_1$  return a span of categories  $F_0A \xleftarrow{\pi_a} \int_{F_1} m \xrightarrow{\pi_b} F_0B$ . The apex category  $\int_{F_1} m$  has for objects pairs  $(\chi, f)$  where  $\chi : \text{disc}X \rightarrow \mathcal{C}$  is an interface and  $f : \chi^{\otimes} \rightarrow I$  is a factor on that interface.  $\chi^{\otimes}$  denotes the monoidal product  $\otimes_{x:X} \chi(x)$  in  $\mathcal{C}$ . The morphisms of  $\int_{F_1} m$  are factorizations of factors through transformations of interfaces; that is to say, a morphism  $(\chi, f) \rightarrow (\chi', f')$  in  $\int_{F_1} m$  is a morphism of interfaces  $\varphi : \chi \rightarrow \chi'$  such that  $f = \varphi^* f'$ , where  $\varphi^*$  denotes precomposition by  $\varphi^{\otimes} := \otimes_{x:X} \varphi_x$ . The functors  $\pi_a$  and  $\pi_b$  forget the factors and project out those parts of the interfaces over  $X$  that are exposed by  $a$  and  $b$  respectively. ( $\int_{F_1} m$  and the functors  $\pi_a, \pi_b$  are in fact obtained through universal constructions — a limit and a Grothendieck construction — as we explain in Appendix B.1. Universality then yields the functorial action of  $F_1$  on morphisms of cospans.)

### 3.3 Factor composition

The functors  $F_0$  and  $F_1$  describe how to attach factors to interfaces, and how these interfaces (and the factors on them) may be transformed by morphisms in  $\mathcal{C}$ . But they do not tell us about the operation at the heart of the open factor graph construction: the horizontal (copy-)composition of factors, or the horizontal composition of their morphisms. These operations are formalized by the external composition structure  $(\mu, \eta)$  with which we now equip  $F$ .

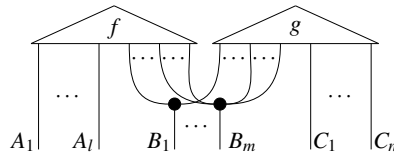
Let us begin with  $\mu$ , a pseudo natural transformation whose component  $\mu_{m,n}$  at a pair of composable cospans  $m = (A \xrightarrow{a} X \xleftarrow{b} B)$  and  $n = (B \xrightarrow{b'} Y \xleftarrow{c} C)$  must be a functor  $\int_{F_1} m \times_{F_0B} \int_{F_1} n \rightarrow \int_{F_1} (n \diamond m)$ . Given factors  $(\chi, f), (\gamma, g)$  in its domain,  $\mu_{m,n}$  must return an interface over  $X +_B Y$  and a factor on that

interface. The former is given by the copairing  $[\chi, \gamma]_B$  of  $\chi$  and  $\gamma$  over  $B$ , while the latter is given by copy-composition. For each  $j : X +_B Y$ , copying induces a morphism

$$\delta_{\chi, \gamma}^j : [\chi, \gamma]_B(j) \rightarrow \bigotimes_{i: [I_B^X, I_B^Y]^{-1}(j)} [\chi, \gamma](i)$$

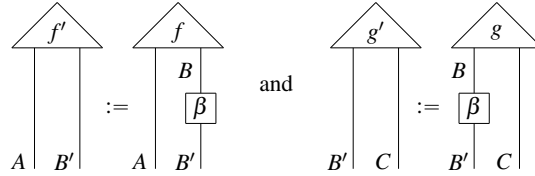
in  $\mathcal{C}$  which is unique up to coassociativity. We then define  $\delta_{\chi, \gamma} := \bigotimes_{j: X +_B Y} \delta_{\chi, \gamma}^j$ , which has the type  $[\chi, \gamma]_B^\otimes \rightarrow [\chi, \gamma]^\otimes$ . We can compose  $f \otimes g$  after this higher-arity copier to yield a factor on the interface  $[\chi, \gamma]_B$ . Thus  $\mu_{m,n}((\chi, f), (\gamma, g)) := ([\chi, \gamma]_B, (f \otimes g) \circ \delta_{\chi, \gamma})$ . (For a more detailed construction and proof that this yields a pseudo natural transformation, see Appendix B.2.)

To see  $\mu$  in action, suppose  $A$  has  $l$ -many elements,  $C$  has  $n$ -many elements, and the coupled interface  $B$  has  $m$ -many elements, with  $B_1$  shared once between  $f$  and  $g$  but  $B_m$  shared an arbitrary number of times.

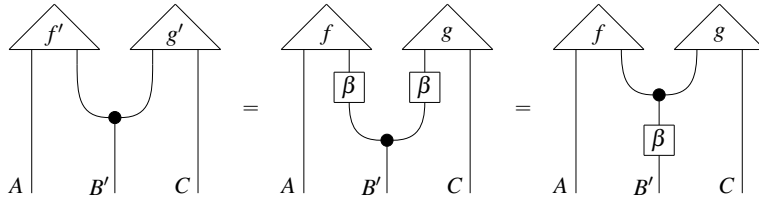


(We have also assumed that  $X \cong A + B$  and  $Y \cong B + C$ , but this need not be the case, as we will see below.)

The naturality of  $\mu_{m,n}$  means that if we were to have  $f'$  and  $g'$  defined by transforming the interface  $B$  along  $\beta$  like so



then copy-composing  $f'$  and  $g'$  along  $B'$  yields the same as copy-composing  $f$  and  $g$  along  $B$  and then transforming the  $B$  interface of the result accordingly:



However, this only works for ‘deterministic’ transformations  $\beta$ : although the rules of open factor graphs do allow us to transform factors using arbitrary morphisms, we can only pull these transformations ‘outside’ the composites when the transformations are comonoid homomorphisms.

It remains to define the unit  $\eta$ , a pseudo natural transformation whose component  $\eta_X$  at each finite set  $X$  is given by a functor  $\eta_X : F_0 X \rightarrow \int F_1(\text{id}_X)$ . We define  $\eta_X$  using the counits of the comonoid structures on the interfaces on  $X$ . That is to say,  $\eta_X$  maps an interface  $\chi : \text{disc } X \rightarrow \mathbf{Comon}(\mathcal{C})$  to the pair  $(\chi, \varepsilon_\chi)$ , where here  $\varepsilon_\chi$  denotes the canonical discarding map  $\chi^\otimes \rightarrow I$ . The functorial action of  $\eta_X$  is obtained from the fact that all morphisms in  $F_0 X$  are comonoid homomorphisms, so if  $\varphi$  is a morphism  $\chi \rightarrow \chi'$  then  $\varepsilon_\chi = \varphi^* \varepsilon_{\chi'}$ . Pseudo naturality of  $\eta$  obtains likewise, and its unitality with respect to  $\mu$  follows from the counitality of the comonoid counits.

### 3.4 Putting it all together

We have defined a double pseudo functor  $F : \mathbf{Cospan}(\mathbf{FinSet}) \rightarrow \mathbf{Span}(\mathbf{Cat})$ , from which we obtain a double category of open factor graphs  $\mathbb{FG}$ , doubly opfibred over  $\mathbf{Cospan}(\mathbf{FinSet})$ , by applying the double Grothendieck construction. This gives  $\mathbb{FG}$  the following structure:

1. Its objects (0-cells) are pairs  $(A, \alpha)$  where  $A$  is a finite set and  $\alpha$  is an object of  $F_0A$  — *i.e.*, an interface in  $\mathcal{C}$  on  $A$ .
2. A vertical 1-cell  $(A, \alpha) \rightarrow (A', \alpha')$  is a pair  $(f, \varphi)$  of a map  $f : A \rightarrow A'$  and a deterministic transformation of interfaces  $\varphi : f_*\alpha \rightarrow \alpha'$  — where  $f_* = F_0(f)$  so that  $\varphi$  is a morphism in  $F_0A'$ .
3. A horizontal 1-cell  $(A, \alpha) \rightarrow (B, \beta)$  is a quintuple  $(X, a, b, \chi, p)$  where  $(A \xrightarrow{a} X \xleftarrow{b} B)$  is a cospan of finite sets,  $\chi$  is an interface on  $X$ , and  $p$  is a factor on  $\chi$  — *i.e.*,  $(\chi, p)$  is an object of  $\int F_1(X, a, b)$  — such that  $\pi_a(\chi, p) = \alpha$  and  $\pi_b(\chi, p) = \beta$ .

4. A 2-cell as in the diagram
 
$$\begin{array}{ccc}
 (A, \alpha) & \xrightarrow{(X, a, b, \chi, p)} & (B, \beta) \\
 (f', \varphi') \downarrow & & \downarrow (f^r, \varphi^r) \\
 (A', \alpha') & \xrightarrow{(X', a', b', \chi', p')} & (B', \beta')
 \end{array}$$
 is a pair  $(f, \varphi)$  where  $f : X \rightarrow X'$

is a map of finite sets and  $\varphi$  is a transformation of interfaces  $f_*\chi \rightarrow \chi'$ , such that  $f_*p = \varphi^*p'$ ,  $\pi_{a'}(\varphi) = \varphi^l$ ,  $\pi_{b'}(\varphi) = \varphi^r$  (with  $\varphi^l$  and  $\varphi^r$  deterministic), and such that  $(f, f^l, f^r)$  constitutes a morphism of cospans, meaning that  $a' \circ f^l = f \circ a$  and  $b' \circ f^r = f \circ b$ .

5. Vertical 1-cells  $(f, \varphi) : (A, \alpha) \rightarrow (A', \alpha')$  and  $(f', \varphi') : (A', \alpha') \rightarrow (A'', \alpha'')$  compose to yield  $(f' \circ f, \varphi' \circ \varphi) : (A, \alpha) \rightarrow (A'', \alpha'')$ ; the vertical composition of 2-cells is similar. The vertical identity 1-cell on  $(A, \alpha)$  is the pair  $(\text{id}_A, \text{id}_\alpha)$  of identity morphisms  $\text{id}_A : A \rightarrow A$  and  $\text{id}_\alpha : \alpha \rightarrow \alpha$ . We denote vertical composition by  $\circ$ .
6. Finally, horizontal 1-cells (and 2-cells, horizontally) compose by cospan- and copy-composition, as described in §3.3. Horizontal identities are given by the external unit  $\eta$ . We denote horizontal composition by  $\diamond$ .

All told, this (plus the details in Appendix B) yields the following theorem.

**Theorem 3.2** (Open factor graphs). Given a symmetric monoidal copy discard category  $(\mathcal{C}, I, \otimes)$ , the foregoing data define a pseudo double category  $\mathbb{FG}$  doubly opfibred over  $\mathbf{Cospan}(\mathbf{FinSet})$ .

**Remark 3.3.** Since each of the elements of the construction of  $\mathbb{FG}$  are functorial, we expect the construction itself to be functorial, so that a functor  $\mathcal{C} \rightarrow \mathcal{C}'$  yields a double functor  $\mathbb{FG}_{\mathcal{C}} \rightarrow \mathbb{FG}_{\mathcal{C}'}$ . But we have not established this, and leave it to future work.

In other future work, we hope to make use of  $\mathbb{FG}$  to study the compositionality of inference on factor graphs, and to study the formal relationships between directed and undirected models.

### 3.5 On a graphical calculus

We motivated the construction of  $\mathbb{FG}$  with the graphic Figure 1. Decorated cospan categories as originally formulated inherit a graphical calculus from their status as hypergraph categories or undirected wiring diagram algebras. Lacking a formal theory of double operads and their graphical calculi, we do not have a similarly formal graphical calculus for  $\mathbb{FG}$ ; still, we expect that its horizontal bicategory (appropriately quotiented) would fit the undirected wiring diagram pattern, at the cost of the transformations of interfaces. Inspired by this, and in an attempt not to pay that cost, in Appendix B.3 we (informally) propose a graphical calculus that matches and extends the non-compositional depictions in the statistics literature.

## 4 References

- [1] Toby St Clere Smithe. “Approximate Inference via Fibrations of Statistical Games”. In: *Electronic Proceedings in Theoretical Computer Science*. Applied Category Theory 2023. Vol. 397. College Park, MD: Open Publishing Association, 06/29/2023, pp. 279–298. DOI: 10.4204/EPTCS.397.17. arXiv: 2306.17009 [math, stat]. URL: <http://arxiv.org/abs/2306.17009> (visited on 09/11/2023).
- [2] Brendan Fong. “Causal Theories: A Categorical Perspective on Bayesian Networks”. University of Oxford, 2013-01-26, 2013. arXiv: 1301.6201v1 [math.PR]. URL: <http://arxiv.org/abs/1301.6201>.
- [3] Kenta Cho and Bart Jacobs. “Disintegration and Bayesian Inversion via String Diagrams”. In: *Math. Struct. Comp. Sci.* 29 (2019) 938–971 (08/29/2017). DOI: 10.1017/S0960129518000488. arXiv: 1709.00322v3 [cs.AI].
- [4] Evan Patterson. *Structured and Decorated Cospans from the Viewpoint of Double Category Theory*. 04/02/2023. arXiv: 2304.00447 [math]. URL: <http://arxiv.org/abs/2304.00447> (visited on 04/20/2023). preprint.
- [5] Sam Staton. “Commutative Semantics for Probabilistic Programming”. In: *Programming Languages and Systems*. Springer Berlin Heidelberg, 2017, pp. 855–879. DOI: 10.1007/978-3-662-54434-1\_32.
- [6] Kenta Cho et al. “An Introduction to Effectus Theory”. 2015. arXiv: 1512.05813 [cs.LO].
- [7] Urs Schreiber. “Quantization via Linear Homotopy Types”. 02/27/2014. arXiv: 1402.7041 [math-ph].
- [8] Brendan Fong. “Decorated Cospans”. In: *Theory and Applications of Categories* 30.33 (2015), pp. 1096–1120. arXiv: 1502.00872. URL: <http://www.tac.mta.ca/tac/volumes/30/33/30-33abs.html>.
- [9] Geoffrey Cruttwell et al. *Double Fibrations*. 05/30/2022. arXiv: 2205.15240 [math]. URL: <http://arxiv.org/abs/2205.15240> (visited on 03/20/2023). preprint.
- [10] Martin J. Wainwright and Michael I. Jordan. “Graphical Models, Exponential Families, and Variational Inference”. In: *Foundations and Trends® in Machine Learning* 1.1–2 (2007), pp. 1–305. ISSN: 1935-8245. DOI: 10.1561/22000000001. URL: <http://dx.doi.org/10.1561/22000000001>.
- [11] Justin Dauwels. “On Variational Message Passing on Factor Graphs”. In: *2007 IEEE International Symposium on Information Theory*. 2007 IEEE International Symposium on Information Theory. Nice: IEEE, 06/2007, pp. 2546–2550. ISBN: 978-1-4244-1397-3. DOI: 10.1109/ISIT.2007.4557602. URL: <http://ieeexplore.ieee.org/document/4557602/> (visited on 11/14/2023).

- [12] Tobias Fritz. “A Synthetic Approach to Markov Kernels, Conditional Independence and Theorems on Sufficient Statistics”. In: *Advances in Mathematics* 370.107239 (08/19/2019). DOI: 10.1016/j.aim.2020.107239. arXiv: 1908.07021v3 [math.ST].
- [13] David I. Spivak. *The Operad of Wiring Diagrams: Formalizing a Graphical Language for Databases, Recursion, and Plug-and-Play Circuits*. 05/01/2013. arXiv: 1305.0297 [cs, math]. URL: <http://arxiv.org/abs/1305.0297> (visited on 12/23/2022). preprint.
- [14] Donald Yau. *Operads of Wiring Diagrams*. Vol. 2192. Lecture Notes in Mathematics. Cham: Springer International Publishing, 2018. ISBN: 978-3-319-95000-6 978-3-319-95001-3. DOI: 10.1007/978-3-319-95001-3. arXiv: 1512.01602 [math]. URL: <http://link.springer.com/10.1007/978-3-319-95001-3> (visited on 09/11/2023).
- [15] Brendan Fong and David I. Spivak. “Hypergraph Categories”. In: *Journal of Pure and Applied Algebra* 223.11 (11/2019), pp. 4746–4777. ISSN: 00224049. DOI: 10.1016/j.jpaa.2019.02.014. arXiv: 1806.08304v3 [math.CT]. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0022404919300489> (visited on 12/17/2022).
- [16] Joe Moeller and Christina Vasilakopoulou. “Monoidal Grothendieck Construction”. In: *Theory and Applications of Categories* 35.31 (2020), pp. 1159–1207. arXiv: 1809.00727v2 [math.CT].
- [17] Magnus Koudahl, Thijs van de Laar, and Bert de Vries. *Realising Synthetic Active Inference Agents, Part I: Epistemic Objectives and Graphical Specification Language*. 06/13/2023. arXiv: 2306.08014 [cs]. URL: <http://arxiv.org/abs/2306.08014> (visited on 07/28/2023). preprint.
- [18] Bob Coecke and Aleks Kissinger. “Categorical Quantum Mechanics II: Classical-Quantum Interaction”. 05/27/2016. arXiv: 1605.08617v1 [quant-ph].

## A Directed models

### A.1 Kernels bifibration

*Proof of Proposition 2.10.* Suppose  $(E, p) \rightsquigarrow^h (F, q) \rightsquigarrow^k (G, r)$  over  $B$ . Then, by the definitions of fibrewise restriction and of kernel composition,

$$(k \circ h)[b] : (x_b, V_b) \mapsto \int_{y:F} k(y, \pi_G(V_b)) h(\pi_E(x_b), dy) .$$

Now, by the condition on kernels over  $B$ , the measure  $h(\pi_E(x_b), -)$  is only supported over  $b$ , so we can write the integral as

$$\int_{y_b:q[b]} k(\pi_F(y_b), \pi_G(V_b)) h(\pi_E(x_b), \pi_F(dy)) .$$

But, again by their definition, this is equal to

$$\int_{y_b:q[b]} k[b](y_b, V_b) h[b](x_b, dy_b)$$

and hence to  $(k[b] \circ h[b])(x_b, V_b)$ . So  $(k \circ h)[b] = k[b] \circ h[b]$ .  $\square$

*Proof of Proposition 2.14.* We need to exhibit an isomorphism

$$\mathcal{K}_B(p, \Delta_f q) \begin{array}{c} \xrightarrow{(-)^b} \\ \xleftarrow{(-)^\sharp} \end{array} \mathcal{K}_C(\Sigma_f p, q)$$

natural in  $p$  and  $q$ .

Suppose then that we have  $\beta : p \rightsquigarrow \Delta_f q$  in  $\mathcal{K}_B$  and  $\gamma : \Sigma_f p \rightsquigarrow q$  in  $\mathcal{K}_C$  as in the following diagrams:

$$\begin{array}{ccc} E & \overset{\beta}{\rightsquigarrow} & f^*F \\ & \searrow p & \swarrow f^*q \\ & & B \end{array} \qquad \begin{array}{ccc} E & \overset{\gamma}{\rightsquigarrow} & F \\ & \searrow p & \swarrow q \\ & & B \\ & \searrow f & \swarrow \\ & & C \end{array}$$

Define  $\beta^b : \Sigma_f p \rightsquigarrow q$  by the kernel

$$E \times \Sigma_F \rightarrow [0, \infty] : (x, V) \mapsto \beta(x, \pi_F^{-1}(V))$$

where  $\pi_F^{-1}(V)$  is the preimage of  $V$  through the projection  $\pi_F : f^*F \rightarrow F$  out of the pullback  $f^*F$ .

Dually, define  $\gamma^\sharp : p \rightsquigarrow \Delta_f q$  by the kernel

$$E \times \Sigma_{f^*F} \rightarrow [0, \infty] : (x, U) \mapsto \gamma(x, \pi_F(U)) .$$

Our first task is to check that these yield well-defined morphisms: that  $\beta^b$  and  $\gamma^\sharp$  satisfy the triangle conditions in  $\mathcal{K}_C$  and  $\mathcal{K}_B$  respectively, that  $\delta_q \circ \beta^b = \delta_{f \circ p}$  and  $\delta_{f^*q} \circ \gamma^\sharp = \delta_p$ . For the former, we have the following chain of equalities:

$$\begin{aligned} (\delta_q \circ \beta^b)(x, V) &= \beta^b(x, q^{-1}(V)) && \text{by pushforward} \\ &= \beta(x, \pi_F^{-1}q^{-1}(V)) && \text{by definition of } b \\ &= \beta(x, (f^*q)^{-1}f^{-1}(V)) && \text{by pullback of } q \text{ along } f \\ &= (\delta_{f^*q} \circ \beta)(x, f^{-1}(V)) && \text{by pushforward} \\ &= [p(x) \in f^{-1}(V)] && \text{since } \delta_p = \delta_{f^*q} \circ \beta \text{ in } \mathcal{K}_B \\ &= [(f \circ p)(x) \in V] && \text{by preimage} \\ &= \delta_{f \circ p}(x, V) && \text{by definition of } \delta \end{aligned}$$

Similarly, for the latter, we have the following:

$$\begin{aligned} (\delta_{f^*q} \circ \gamma^\sharp)(x, U) &= \gamma^\sharp(x, (f^*q)^{-1}(U)) && \text{by pushforward} \\ &= \gamma(x, \pi_F((f^*q)^{-1}(U))) && \text{by definition of } \sharp \\ &= \gamma(x, q^{-1}(f(U))) && \text{by pullback of } q \text{ along } f \\ &= (\delta_q \circ \gamma)(x, q^{-1}(f(U))) && \text{by pushforward} \\ &= [p(x) \in f^{-1}(f(U))] && \text{since } \delta_{f \circ p} = \delta_q \circ \gamma \text{ in } \mathcal{K}_C \\ &= [p(x) \in U] && \text{by preimage} \\ &= \delta_p(x, U) && \text{by definition of } \delta \end{aligned}$$

This establishes that  $\beta^b$  and  $\gamma^\sharp$  are well-defined morphisms in their respective fibre categories. That the maps  $(-)^b$  and  $(-)^{\sharp}$  constitute an isomorphism of hom sets is immediate from the fact that images and preimages are mutually inverse:  $\pi_F^{-1}(\pi_F(U)) = U$  and  $\pi_F(\pi_F^{-1}(V)) = V$ .

It therefore only remains to demonstrate naturality. Since  $(-)^b$  and  $(-)^{\sharp}$  are mutually inverse, we need only demonstrate naturality for one; we choose  $(-)^{\sharp}$ . This means showing that the following square commutes for all kernels  $\varphi, \psi$  of appropriate type:

$$\begin{array}{ccc} \mathcal{K}_C(\Sigma_f p, q) & \xrightarrow{(-)^{\sharp}_{p,q}} & \mathcal{K}_B(p, \Delta_f q) \\ \mathcal{K}_C(\Sigma_f \varphi, \psi) \downarrow & & \downarrow \mathcal{K}_B(\varphi, \Delta_f \psi) \\ \mathcal{K}_C(\Sigma_f o, r) & \xrightarrow{(-)^{\sharp}_{o,r}} & \mathcal{K}_B(o, \Delta_f r) \end{array}$$

Since the hom functors  $\mathcal{K}_C$  and  $\mathcal{K}_B$  act by pre- and post-composition, this requires establishing that

$$o \overset{\varphi}{\rightsquigarrow} p \overset{\gamma^\sharp}{\rightsquigarrow} \Delta_f q \overset{\Delta_f \psi}{\rightsquigarrow} \Delta_f r = \left( \Sigma_f o \overset{\Sigma_f \varphi}{\rightsquigarrow} \Sigma_f p \overset{\gamma}{\rightsquigarrow} q \overset{\psi}{\rightsquigarrow} r \right)^\sharp.$$

We noted above that  $\Sigma_f p[b] \cong p[f^*b]$ ; likewise, we have  $\Delta_f q[b] \cong q[f \circ b]$ , since the pasting of two pullback squares is again a pullback square:

$$\begin{array}{ccccc} (\Delta_f q)[b] \cong q[f \circ b] & \xrightarrow{\pi_{f^*F}} & f^*F & \xrightarrow{\pi_F} & F \\ (f \circ b)^*q \downarrow & \lrcorner & f^*q \downarrow & \lrcorner & \downarrow q \\ J & \xrightarrow{b} & B & \xrightarrow{f} & C \end{array}$$

Thus, working fibrewise, we have

$$\gamma^\sharp[b](x, U) = \gamma^\sharp(\pi_E(x), \pi_{f^*F}(U)) = \gamma(\pi_E(x), (\pi_F \circ \pi_{f^*F})(U)) = \gamma[f \circ b](x, U)$$



i.e.  $\gamma_b^\# = \gamma[f \circ b]$ , and so

$$\begin{aligned}
(\psi \circ \gamma \circ \Sigma_f \varphi)^\# [b](w, V) &= (\psi \circ \gamma \circ \Sigma_f \varphi)[f \circ b](w, V) \\
&\quad (\text{since } \gamma^\# [b] = \gamma[f \circ b]) \\
&= \int_{y:q[f \circ b]} \psi[f \circ b](y, V) \int_{x:\Sigma_f p[f \circ b]} \gamma[f \circ b](x, dy) (\Sigma_f \varphi)[f \circ b](w, dx) \\
&\quad (\text{by fibrewise composition}) \\
&= \int_{y:q[f \circ b]} \psi[f \circ b](y, V) \int_{x:p[b]} \gamma[f \circ b](x, dy) \varphi[b](w, dx) \\
&\quad (\text{since } \Sigma_f p[f \circ b] \cong p[f^*(f \circ b)] \cong p[b]) \\
&= \int_{y:q[f \circ b]} \psi[f \circ b](y, V) \int_{x:p[b]} \gamma^\# [b](x, dy) \varphi[b](w, dx) \\
&\quad (\text{since } \gamma^\# [b] = \gamma[f \circ b]) \\
&= \int_{y:\Delta_f q[b]} (\Delta_f \psi)[b](y, V) \int_{x:p[b]} \gamma^\# [b](x, dy) \varphi[b](w, dx) \\
&\quad (\text{since } q[f \circ b] \cong \Delta_f q[b]) \\
&= (\Delta_f \psi \circ \gamma^\# \circ \varphi)[b](w, V) \\
&\quad (\text{by fibrewise composition}).
\end{aligned}$$

This establishes that  $\Sigma_f \dashv \Delta_f$ . □

*Proof of Proposition 2.15.* The functor  $\Sigma_\rho \Delta_\pi : \mathcal{K}_E \rightarrow \mathcal{K}_F$  sends  $X \xrightarrow{\alpha} E$  to  $\pi^* X \xrightarrow{\Delta_\pi \alpha} P \xrightarrow{\rho} F$ ; fibrewise, by adjointness, we have  $\Sigma_\rho \Delta_\pi(\alpha)[i] \cong \alpha[\pi \circ \rho^* i]$  and, on morphisms  $f$ , we have  $\Sigma_\rho \Delta_\pi(f)[i] = f[\pi \circ \rho^* i]$ . On the other side, the functor  $\Delta_q \Sigma_p$  sends  $\alpha$  to  $q^* X \xrightarrow{\Delta_q \alpha} q^* E \xrightarrow{\Delta_q \rho} F$ ; fibrewise, we have  $\Delta_q \Sigma_p(\alpha)[i] \cong \alpha[p^*(q \circ i)]$  on objects and  $\Delta_q \Sigma_p(f)[i] = f[p^*(q \circ i)]$  on morphisms. It therefore suffices to show that  $p^*(q \circ i) = \pi \circ \rho^* i$  in **Meas**. This follows from the universal property of the pullback, which implies that the pasting of two pullback squares is again a pullback square:

$$\begin{array}{ccccc}
& & p^*(q \circ i) & & \\
& \nearrow & & \searrow & \\
\bullet & \xrightarrow{\rho^* i} & P & \xrightarrow{\pi} & E \\
& \downarrow & \rho \downarrow & & \downarrow p \\
I & \xrightarrow{i} & F & \xrightarrow{q} & B
\end{array}$$

The outer and inner squares are thus both pullbacks of  $p$  along  $q \circ i$ , and hence we must have  $p^*(q \circ i) = \pi \circ \rho^* i$ . This yields an isomorphism  $\Sigma_\rho \Delta_\pi(\alpha) \cong \Delta_q \Sigma_p(\alpha)$  rather than an equality because universal constructions are only unique up to natural isomorphism; and thus the naturality of this isomorphism follows similarly from universality. □

## A.2 Pull-push sections

*Proof of Theorem 2.17.* Much of the structure is simplified, owing to the trivial vertical decorations and the decoration of the apices by sets. Thus, for example, each component  $\mu_{m,n}$  or  $\eta_E$  of the pseudo natural

transformations  $\mu$  or  $\eta$  is only a function (discrete functor). In the main text, the functoriality of  $S_0$  and  $S_1$  is justified. So here we will only justify the (pseudo) naturality of  $\mu$  and  $\eta$  and their (pseudo) associativity and unitality.

Suppose then that  $f : m \rightarrow m'$  and  $g : n \rightarrow n'$  are horizontally composable morphisms of spans (i.e., such that  $g_l = f_r$ ) from  $m = (A \xleftarrow{a} E \xrightarrow{b} B)$  and  $n = (B \xleftarrow{\bar{b}} F \xrightarrow{c} C)$  to  $m' = (A' \xleftarrow{a'} E' \xrightarrow{b'} B')$  and  $n' = (B' \xleftarrow{\bar{b}'} F' \xrightarrow{c'} C')$ . To validate the pseudo naturality of  $\mu$  is to confirm the commutativity of the square

$$\begin{array}{ccc} S_1(m') \times S_1(n') & \xrightarrow{\mu_{m',n'}} & S_1(n' \diamond m') \\ S_1(f) \times S_1(g) \downarrow & & \downarrow S_1(g \diamond f) \\ S_1(m) \times S_1(n) & \xrightarrow{\mu_{m,n}} & S_1(n \diamond m) \end{array} \cdot$$

up to natural isomorphism. The top-right composite evaluates to

$$\begin{aligned} S_1(g \diamond f)(\Sigma_{a'} \Delta_{b'}(\tau') \circ \sigma') &= \Delta_{f_l}(\Sigma_{a'} \Delta_{b'}(\tau') \circ \sigma') && \text{(by definition)} \\ &= \Delta_{f_l} \Sigma_{a'} \Delta_{b'}(\tau') \circ \Delta_{f_l}(\sigma') && \text{(by functoriality)} \end{aligned}$$

while the left-bottom composite evaluates to

$$\begin{aligned} \Sigma_a \Delta_b(S_1(g)(\tau') \circ S_1(f)(\sigma')) &= \Sigma_a \Delta_b \Delta_{g_l}(\tau') \circ \Delta_{f_l}(\sigma') && \text{(by definition)} \\ &= \Sigma_a \Delta_f \Delta_{b'}(\tau') \circ \Delta_{f_l}(\sigma') && \text{(since } g_l \circ b = b' \circ f) \\ &\cong \Delta_{f_l} \Sigma_{a'} \Delta_{b'}(\tau') \circ \Delta_{f_l}(\sigma') && \text{(by Beck-Chevalley)} \end{aligned}$$

and so  $S_1(g \diamond f) \circ \mu_{m',n'} \cong \mu_{m,n} \circ (S_1(f) \times S_1(g))$  as required.

Validating the pseudo naturality of  $\eta$  means, for any  $f : E \rightarrow F$  in  $\mathcal{B}$ , validating that  $\eta_F \cong S_1(f) \circ \eta_E$ , i.e.,  $\eta_F \cong \Delta_f \eta_E$ . Since  $\eta_E$  is the identity on  $\iota E$  and  $\Delta_f$  is a functor, it must preserve identities. So in fact  $\eta$  is strictly natural (just as pull-push copy-composition is strictly unital, as we confirm below).

To validate pseudo associativity, suppose  $m$  and  $n$  are spans as before and  $o = (C \xleftarrow{\bar{c}} G \xrightarrow{d} D)$ . We need to show that  $\mu_{n \diamond m, o}(\mu_{m,n}(\sigma, \tau), \rho) \cong \mu_{m, o \diamond n}(\sigma, \mu_{n,o}(\tau, \rho))$  for all appropriately typed sections  $\sigma, \tau, \rho$ . This obtains as follows:

$$\begin{aligned} \mu_{n \diamond m, o}(\mu_{m,n}(\sigma, \tau), \rho) &= \mu_{n \diamond m, o}(\Sigma_a \Delta_b(\tau) \circ \sigma, \rho) && \text{(by definition)} \\ &= \Sigma_a \Sigma_{\pi_E} \Delta_{\pi_F} \Delta_c(\rho) \circ \Sigma_a \Delta_b(\tau) \circ \sigma && \text{(by definition)} \\ &\cong \Sigma_a \Delta_b \Sigma_{\bar{b}} \Delta_c(\rho) \circ \Sigma_a \Delta_b(\tau) \circ \sigma && \text{(by Beck-Chevalley)} \\ &= \Sigma_a \Delta_b(\Sigma_{\bar{b}} \Delta_c(\rho) \circ \tau) \circ \sigma && \text{(by functoriality)} \\ &= \mu_{m, o \diamond n}(\sigma, \Sigma_{\bar{b}} \Delta_c(\rho) \circ \tau) && \text{(by definition)} \\ &= \mu_{m, o \diamond n}(\sigma, \mu_{n,o}(\tau, \rho)) && \text{(by definition)} \end{aligned}$$

where  $\pi_E$  and  $\pi_F$  are the projections out of the pullback of  $\bar{b}$  along  $b$ .

Finally, strict unitality demands that  $\mu_{\text{id}_B, n}(\eta_B, \tau) = \tau$  and  $\mu_{m, \text{id}_B}(\sigma, \eta_B) = \sigma$ . We have

$$\mu_{\text{id}_B, n}(\eta_B, \tau) = \Sigma_{\text{id}_B} \Delta_{\text{id}_B}(\tau) \circ \eta_B = \tau \circ \text{id}_{\iota_B} = \tau$$

and

$$\mu_{m, \text{id}_B}(\sigma, \eta_B) = \Sigma_a \Delta_b(\eta_B) \circ \sigma = \Sigma_a \Delta_b(\text{id}_{\iota_B}) \circ \sigma = \Sigma_a(\text{id}_{\iota_E}) \circ \sigma = \text{id}_{\Sigma_a \iota_E} \circ \sigma = \sigma$$

as required.  $\square$

## B Factor graphs

### B.1 Constructing the horizontal decorations

The apex decoration has a peculiar form, with one category of decorating data (the interfaces) further decorated by a family of others (the factors). This is because the apex decorations are themselves obtained by a Grothendieck construction, defined functorially for each finite set. The first step in this construction is to define the interface categories. Since we have the cospans determining which parts of the interfaces are exposed for factor composition, there may now be some parts which remain ‘latent’, and transformations of these parts are not constrained to be comonoid homomorphisms. That is, we can allow the transformations of the latent parts of the interfaces to be arbitrary morphisms in  $\mathcal{C}$ , as long as the exposed parts remain constrained as before<sup>4</sup>.

This line of thought formalizes as follows. Given a cospan  $m : (A \xrightarrow{a} X \xleftarrow{b} B)$ , we first define an enlarged version of  $F_0X$ , not totally restricted to comonoid homomorphisms. We denote this enlargement by  $\tilde{F}_0m$  and construct it as the following limit of categories, so that we do still have comonoid homomorphisms over the legs:

$$\begin{array}{ccccc}
 \mathbf{Cat}(\text{disc}A, \mathbf{Comon}(\mathcal{C})) & \xleftarrow{\tilde{F}_0m} & & \xrightarrow{\tilde{F}_0m} & \mathbf{Cat}(\text{disc}B, \mathbf{Comon}(\mathcal{C})) \\
 \downarrow & & & & \downarrow \\
 \mathbf{Cat}(\text{disc}A, \mathcal{C}) & \xleftarrow{\mathbf{Cat}(\text{disc}a, \mathcal{C})} & \mathbf{Cat}(\text{disc}X, \mathcal{C}) & \xrightarrow{\mathbf{Cat}(\text{disc}b, \mathcal{C})} & \mathbf{Cat}(\text{disc}B, \mathcal{C})
 \end{array}$$

Note that, because the left and right vertical functors in this diagram are embeddings, so is the induced functor  $\tilde{F}_0m \rightarrow \mathbf{Cat}(\text{disc}X, \mathcal{C})$ . We can therefore understand  $\tilde{F}_0m$  as the wide subcategory of  $\mathbf{Cat}(\text{disc}X, \mathcal{C})$  whose morphisms are those  $X$ -indexed families of morphisms of  $\mathcal{C}$  whose  $A$ - and  $B$ -components are comonoid homomorphisms.

The next step is to decorate  $\tilde{F}_0m$  with factors. Since  $\mathcal{C}$  is a category, factors on an interface  $\chi : \text{disc}X \rightarrow \mathcal{C}$  form a set  $\mathcal{C}(\chi^\otimes, I)$ . We thus define a presheaf  $\underline{F}_1m : \tilde{F}_0m^{\text{op}} \rightarrow \mathbf{Set}$  as follows. On objects (interfaces)  $\chi$ , we define  $\underline{F}_1m(\chi)$  to be  $\mathcal{C}(\chi^\otimes, I)$ . On morphisms of interfaces  $\varphi : \chi \rightarrow \chi'$ , we set  $\underline{F}_1m(\varphi) : \underline{F}_1m(\chi') \rightarrow \underline{F}_1m(\chi)$  to be the precomposition functor  $\mathcal{C}(\varphi^\otimes, I) : \mathcal{C}(\chi'^\otimes, I) \rightarrow \mathcal{C}(\chi^\otimes, I)$  and denote it by  $\varphi^*$ . This yields a well-defined presheaf by the functoriality of  $\otimes$  and of precomposition. The apex decoration returned by  $F_1$  on  $m$  is then given by the 1-categorical Grothendieck construction of  $\underline{F}_1m$  (its category of elements), denoted  $\int \underline{F}_1m$ , whose objects and morphisms are as described in the main text.

This tells us how to decorate the apices of cospans — but, given the cospan  $m$ ,  $F_1$  needs to return a whole span of categories  $F_0A \xleftarrow{\pi_a} \int \underline{F}_1m \xrightarrow{\pi_b} F_0B$ . We obtain the legs of this span through the universal properties of the construction of  $\int \underline{F}_1m$ . First, because  $\int \underline{F}_1m$  is obtained by a Grothendieck construction, it is equipped with a projection functor  $\pi_m : \int \underline{F}_1m \rightarrow \tilde{F}_0m$  which acts by ‘forgetting’ the factors; *i.e.*  $\pi_m(\chi, f) = \chi$ . Next, the construction of  $\tilde{F}_0m$  as a limit means that it is equipped with canonical projection functors to  $F_0A$  and  $F_0B$ :

$$F_0A = \mathbf{Cat}(\text{disc}A, \mathbf{Comon}(\mathcal{C})) \xleftarrow{\pi_a} \int \underline{F}_1m \xrightarrow{\pi_b} \mathbf{Cat}(\text{disc}B, \mathbf{Comon}(\mathcal{C})) = F_0B$$

<sup>4</sup>It is this part of the construction that means we cannot use Fong’s original decorated cospans: here, the decoration on the apex necessarily depends on the legs, in order that we have comonoid homomorphisms in the right places.

Therefore, we obtain the requisite span of categories  $F_1 m := (F_0 A \xleftarrow{\pi_a} \int \underline{F_1} m \xrightarrow{\pi_b} F_0 B)$  as the composition of this span of limit projections with  $\pi_m$ :

$$\begin{array}{ccc}
 & \int \underline{F_1} m & \\
 \pi_a \swarrow & \downarrow \pi_m & \searrow \pi_b \\
 F_0 A & \xleftarrow{\quad} \tilde{F}_0 m \xrightarrow{\quad} & F_0 B
 \end{array}$$

We now need to define the action of  $F_1$  on morphisms of cospans. Suppose that  $m' := (A' \xleftarrow{a'} X' \xrightarrow{b'} B')$  is another cospan. A morphism  $\phi : m \rightarrow m'$  is a triple  $(\phi_l, \phi, \phi_r)$  of maps of finite sets making the following diagram commute:

$$\begin{array}{ccccc}
 A & \xrightarrow{a} & X & \xleftarrow{b} & B \\
 \phi_l \downarrow & & \downarrow \phi & & \downarrow \phi_r \\
 A' & \xrightarrow{a'} & X' & \xleftarrow{b'} & B'
 \end{array}$$

We know from the treatment above that  $F_0$  is functorial, and this yields the action of  $F_1$  on the left and right components  $\phi_l$  and  $\phi_r$ . Moreover, the construction of  $\int \underline{F_1} m$  is functorial: first, the construction of  $\tilde{F}_0 m$  is functorial in  $m$ , by the functoriality of taking limits. Likewise, the Grothendieck construction is functorial, yielding a functor  $\int \underline{F_1} : \mathbf{FinSet}^{\{\bullet \rightarrow \bullet \leftarrow \bullet\}} \rightarrow \mathbf{Cat}$ . The following diagram then commutes by construction:

$$\begin{array}{ccccc}
 F_0 A & \xleftarrow{\pi_a} & \int \underline{F_1} m & \xrightarrow{\pi_b} & F_0 B \\
 F_0 \phi_l \downarrow & & \downarrow \int \underline{F_1} \phi & & \downarrow F_0 \phi_r \\
 F_0 A' & \xleftarrow{\pi_{a'}} & \int \underline{F_1} m' & \xrightarrow{\pi_{b'}} & F_0 B'
 \end{array}$$

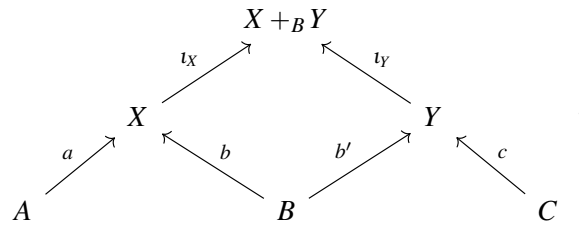
Consequently, we define the action of  $F_1$  on the morphism of cospans  $\phi$  to return the morphism of spans  $F_1 \phi$  whose components are  $(F_0 \phi_l, \int \underline{F_1} \phi, F_0 \phi_r)$ , and we see that this action is functorial by construction.

## B.2 The horizontal composition

$\mu$  must be a pseudo natural transformation filling the following diagram:

$$\begin{array}{ccc}
 \mathbf{FinSet}^{\{\bullet \rightarrow \bullet \leftarrow \bullet\}} \times_{\mathbf{FinSet}} \mathbf{FinSet}^{\{\bullet \rightarrow \bullet \leftarrow \bullet\}} & \xrightarrow{\diamond_{\mathbf{Cospans}(\mathbf{FinSet})}} & \mathbf{FinSet}^{\{\bullet \rightarrow \bullet \leftarrow \bullet\}} \\
 F_1 \times F_1 \downarrow & \nearrow \mu & \downarrow F_1 \\
 \mathbf{Cat}^{\{\bullet \leftarrow \bullet \rightarrow \bullet\}} \times_{\mathbf{Cat}} \mathbf{Cat}^{\{\bullet \leftarrow \bullet \rightarrow \bullet\}} & \xrightarrow{\diamond_{\mathbf{Span}(\mathbf{Cat})}} & \mathbf{Cat}^{\{\bullet \leftarrow \bullet \rightarrow \bullet\}}
 \end{array}$$

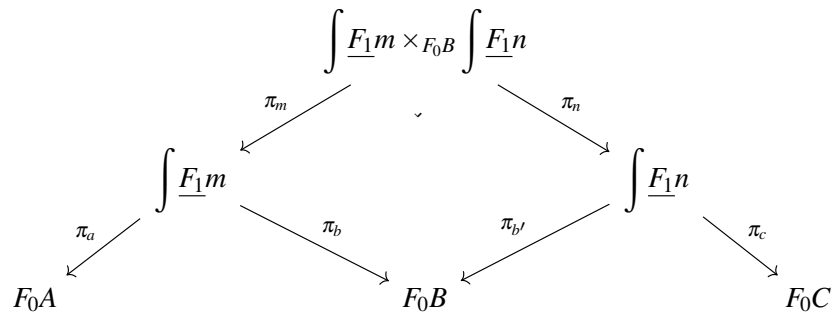
The external composition  $\diamond_{\text{Cospan}(\mathbf{FinSet})}$  acts on composable cospans by pushout, taking  $m : (A \xrightarrow{a} X \xleftarrow{b} B)$  and  $n : (B \xrightarrow{b'} Y \xleftarrow{c} C)$  to the outer cospan in the pushout diagram



(Pseudo functoriality of  $\diamond_{\text{Cospan}(\mathbf{FinSet})}$  follows from the universal properties of colimits.) The composite pseudo functor along the top and right of the above diagram therefore acts on  $m$  and  $n$  to yield the span of categories

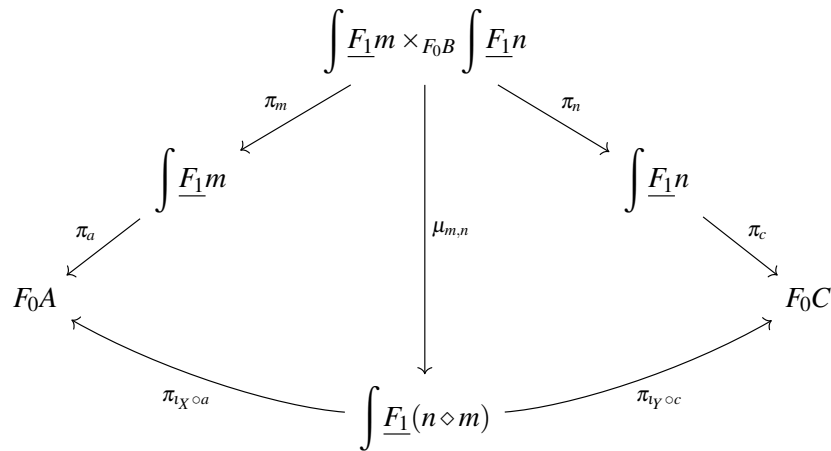
$$F_0A \xleftarrow{\pi_{i_X \circ a}} \int \underline{F}_1(n \diamond m) \xrightarrow{\pi_{i_Y \circ c}} F_0C .$$

Dually,  $\diamond_{\text{Span}(\mathbf{Cat})}$  acts on composable spans by pullback, so that the composite pseudo functor along the left and bottom of the diagram acts to yield the outer span in the following diagram:



Note that, here and throughout this section,  $\pi_m$  and  $\pi_n$  are the projections out of the pullback, rather than the Grothendieck fibrations of the previous section.

In both cases, the feet of the span are  $F_0A$  and  $F_0C$  on the left and right respectively; this must be the case for  $F$  to be a well-defined double functor. Consequently, the component of  $\mu$  at  $(m, n)$ , denoted  $\mu_{m,n}$ , must be a functor  $\int \underline{F}_1 m \times_{F_0 B} \int \underline{F}_1 n \rightarrow \int \underline{F}_1(n \diamond m)$  making this diagram commute:



This means that  $\mu_{m,n}$  takes a pair of factors over  $X$  and  $Y$ , that are composable by agreeing over  $B$ , and returns a composite factor over  $X +_B Y$ , all while preserving the exposed interfaces over  $A$  and  $C$ . We define  $\mu_{m,n}$  to act by copy-composition, as follows.

First, notice that both  $\int \underline{F}_1(n \diamond m)$  and  $\int \underline{F}_1 m \times_{F_0 B} \int \underline{F}_1 n$  are (discrete) fibrations; the former over  $\tilde{F}_0(n \diamond m)$  by definition and the latter over  $\tilde{F}_0 m \times_{F_0 B} \tilde{F}_0 n$ , by the universal property of the pullback:

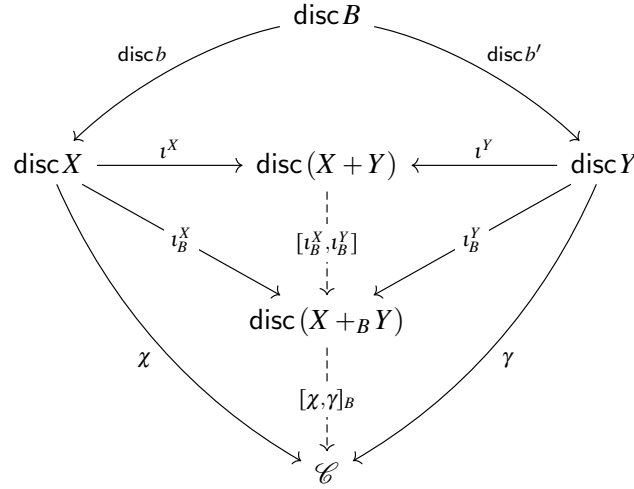
$$\begin{array}{ccccc}
 & & \int \underline{F}_1 m \times_{F_0 B} \int \underline{F}_1 n & & \\
 & \swarrow & \downarrow \text{dashed} & \searrow & \\
 \int \underline{F}_1 m & & & & \int \underline{F}_1 n \\
 \downarrow \pi_m & & \tilde{F}_0 m \times_{F_0 B} \tilde{F}_0 n & & \downarrow \pi_n \\
 \tilde{F}_0 m & & \downarrow \text{dashed} & & \tilde{F}_0 n \\
 & \swarrow & & \searrow & \\
 & & F_0 B & & 
 \end{array}$$

$\mu_{m,n}$  should therefore correspond to a morphism of fibrations. By the Grothendieck construction [16, p.7], such a morphism is equivalent to a pair  $(\mu_{m,n}^0, \mu_{m,n}^1)$  of a functor and a natural transformation as in the following diagram:

$$\begin{array}{ccc}
 (\tilde{F}_0 m \times_{F_0 B} \tilde{F}_0 n)^{\text{op}} & \xrightarrow{\mu_{m,n}^{0 \text{ op}}} & \tilde{F}_0(n \diamond m)^{\text{op}} \\
 \downarrow & \searrow & \downarrow \\
 (\underline{F}_1 m, \underline{F}_1 n) & \xrightarrow{\mu_{m,n}^1} & \underline{F}_1(n \diamond m) \\
 & \searrow & \\
 & & \mathbf{Set}
 \end{array}$$

To define the functor  $\mu_{m,n}^0 : \tilde{F}_0 m \times_{F_0 B} \tilde{F}_0 n \rightarrow \tilde{F}_0(n \diamond m)$ , note that the objects in its domain are pairs  $(\chi, \gamma)$  of functors  $\chi : \text{disc} X \rightarrow \mathcal{C}$  and  $\gamma : \text{disc} Y \rightarrow \mathcal{C}$  that agree on  $B$ . By the universal properties of the coproduct  $X + Y$  and the pushout  $X +_B Y$ , there must be copairings  $[\chi, \gamma] : \text{disc}(X + Y) \rightarrow \mathcal{C}$  and  $[\chi, \gamma]_B : \text{disc}(X +_B Y) \rightarrow \mathcal{C}$  making the following diagram commute; in particular,  $[\chi, \gamma]$  must factor

through  $[\chi, \gamma]_B$ :



Since the image of  $(\chi, \gamma)$  under  $\mu_{m,n}^0$  must be a functor  $\text{disc}(X +_B Y) \rightarrow \mathcal{C}$ , we make the universal definition  $\mu_{m,n}^0(\chi, \gamma) := [\chi, \gamma]_B$ . Functoriality follows from the functoriality of colimits.

The corresponding component of  $\mu_{m,n}^1$  must then be a function

$$\mu_{m,n}^1(\chi, \gamma) : \mathcal{C}(\chi^\otimes, I) \times \mathcal{C}(\gamma^\otimes, I) \rightarrow \mathcal{C}([\chi, \gamma]_B^\otimes, I).$$

Note that  $[\chi, \gamma]^\otimes \cong \chi^\otimes \otimes \gamma^\otimes$ . We can thus factor  $\mu_{m,n}^1$  as

$$\mathcal{C}(\chi^\otimes, I) \times \mathcal{C}(\gamma^\otimes, I) \xrightarrow{\otimes} \mathcal{C}(\chi^\otimes \otimes \gamma^\otimes, I) \xrightarrow{\sim} \mathcal{C}([\chi, \gamma]^\otimes, I) \rightarrow \mathcal{C}([\chi, \gamma]_B^\otimes, I)$$

so that, by Yoneda, we seek a morphism  $\delta_{\chi, \gamma} : [\chi, \gamma]_B^\otimes \rightarrow [\chi, \gamma]^\otimes$  in  $\mathcal{C}$ .

It is here that copying finally enters. For each  $j : X +_B Y$ , copying induces a morphism

$$\delta_{\chi, \gamma}^j : [\chi, \gamma]_B(j) \rightarrow \bigotimes_{i: [\iota_B^X, \iota_B^Y]^{-1}(j)} [\chi, \gamma](i)$$

in  $\mathcal{C}$  which is unique up to coassociativity. Note that the codomain of this morphism is, by the pullback constraint, the monoidal product of  $[\iota_B^X, \iota_B^Y]^{-1}(j)$ -many copies of the same object  $[\chi, \gamma]_B(j)$  so that this copying is well-defined; and when  $j$  is not in the image of  $B$ , then this product is only unary, so that in this case,  $\delta_{\chi, \gamma}^j$  is the corresponding identity morphism. Then

$$\bigotimes_{j: X+B Y} \delta_{\chi, \gamma}^j : \bigotimes_{j: X+B Y} [\chi, \gamma]_B(j) \rightarrow \bigotimes_{j: X+B Y} \bigotimes_{i: [\iota_B^X, \iota_B^Y]^{-1}(j)} [\chi, \gamma](i)$$

has the requisite type, since  $\bigotimes_{j: X+B Y} [\chi, \gamma]_B(j) = [\chi, \gamma]_B^\otimes$  and

$$\bigotimes_{j: X+B Y} \bigotimes_{i: [\iota_B^X, \iota_B^Y]^{-1}(j)} [\chi, \gamma](i) = \bigotimes_{i: X+Y} [\chi, \gamma](i) = [\chi, \gamma]^\otimes$$

both by definition. Hence we define  $\delta_{\chi, \gamma} := \bigotimes_{j: X+B Y} \delta_{\chi, \gamma}^j$ .

We need to verify that  $\delta_{\chi, \gamma}$  is natural in  $\chi$  and  $\gamma$ . Since identity morphisms are always natural, we only need to check the case that  $[\iota_B^X, \iota_B^Y]^{-1}(j)$  has more than one element — and this in turn only obtains for  $j$  in

the image of  $B$ . By construction, morphisms of interfaces on those components are necessarily comonoid homomorphisms, and these are the morphisms for which  $\delta_{\chi,\gamma}^j$  is natural. Therefore,  $\delta_{\chi,\gamma}$  is always natural in  $\chi$  and  $\gamma$ , and since every factor of  $\mu_{m,n}^1$  is accordingly natural, so is  $\mu_{m,n}^1$  itself, as required.

This means that  $(\mu_{m,n}^0, \mu_{m,n}^1)$  constitutes a well-defined morphism of indexed sets, thereby inducing a well-defined morphism  $\mu_{m,n}$  of discrete fibrations. Thus, given factors  $(\chi, f)$  and  $(\gamma, g)$  that agree over  $B$ , the external composition  $\mu_{m,n}$  acts to return  $([\chi, \gamma]_B, (f \otimes g) \circ \delta_{\chi,\gamma})$ .

We also need to define the unit  $\eta$ , a pseudo natural transformation as in the following diagram:

$$\begin{array}{ccc}
 \mathbf{FinSet} & \xrightarrow{\text{id}_{\text{Cospan}(\mathbf{FinSet})}} & \mathbf{FinSet}^{\{\bullet \rightarrow \bullet \leftarrow \bullet\}} \\
 \downarrow F_0 & \nearrow \eta & \downarrow F_1 \\
 \mathbf{Cat} & \xrightarrow{\text{id}_{\text{Span}(\mathbf{Cat})}} & \mathbf{Cat}^{\{\bullet \leftarrow \bullet \rightarrow \bullet\}}
 \end{array}$$

The component  $\eta_X$  at each finite set  $X$  is determined by a functor  $\eta_X : F_0X \rightarrow \int F_1(\text{id}_X)$  as in the diagram

$$\begin{array}{ccccc}
 F_0X & \xlongequal{\quad} & F_0X & \xlongequal{\quad} & F_0X \\
 \parallel & & \downarrow \eta_X & & \parallel \\
 F_0X & \xleftarrow{\pi_X} & \int F_1(\text{id}_X) & \xrightarrow{\pi_X} & F_0X
 \end{array}$$

where  $\text{id}_X$  denotes the identity cospan  $X = X = X$  on  $X$ . Note that because  $\int F_1(\text{id}_X)$  is defined over the identity span,  $\tilde{F}_0(\text{id}_X) \cong F_0X$ , and so  $\pi_X$  simply forgets the factor decorations. Moreover, this means that all components of morphisms of interfaces in  $\tilde{F}_0(\text{id}_X)$  are comonoid homomorphisms in  $\mathcal{C}$ .

We define  $\eta_X$  using the counits of the comonoid structures on the interfaces on  $X$ . That is to say,  $\eta_X$  maps an interface  $\chi : \text{disc}X \rightarrow \mathbf{Comon}(\mathcal{C})$  to the pair  $(\chi, \varepsilon_\chi)$ , where here  $\varepsilon_\chi$  denotes the canonical discarding map  $\chi^\otimes \rightarrow I$ . The functoriality of  $\eta_X$  follows from the fact that all morphisms in  $F_0X$  are comonoid homomorphisms, so if  $\varphi$  is a morphism  $\chi \rightarrow \chi'$  then  $\varepsilon_\chi = \varphi^* \varepsilon_{\chi'}$ . Pseudo naturality of  $\eta$  obtains likewise, and its unitality with respect to  $\mu$  follows from the counitality of the comonoid counits.

At this final point, we should also validate the remaining laws that a lax double functor should satisfy: in particular, the associativity of  $\mu$ . However, we shall at this point simply assert that associativity follows from the associativity of  $\otimes$ , the coassociativity of the relevant comonoid structures, and the universal properties of the (co)limits involved in the constructions, leaving a detailed examination of this claim to future exposition.

### B.3 Graphical calculus

The relationship between decorated cospans and undirected wiring diagram algebras indicates that factor graphs should be understood as composing ‘operadically’ (multicategorically): that is to say, by nesting; thus, within each factor of a factor graph may be hiding a whole other factor graph. This formalizes the situation sometimes encountered in the literature — *e.g.* Koudahl, van de Laar, and de Vries [17, Fig.20] — in which a collection of factors is grouped into a composite factor, sometimes depicted by drawing an extra box around the collection.

To hew more closely to existing undirected wiring diagrams, as well as the standard string diagrams for monoidal categories, we will adopt a syntax in which factors are depicted with circular boxes, exposed variables by emanating wires, and transformation morphisms by rectangular boxes. We will



depict comonoid homomorphisms as ‘diamond’ shaped boxes, both to indicate that they commute with horizontal composition  $\diamond$  and also to suggest that they can pierce through the bubble enclosing a factor; this means we will attach wires to their corners. Conversely, we will depict arbitrary morphisms with standard rectangular boxes, attaching wires to their sides.

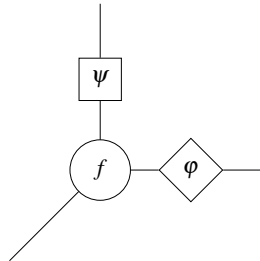
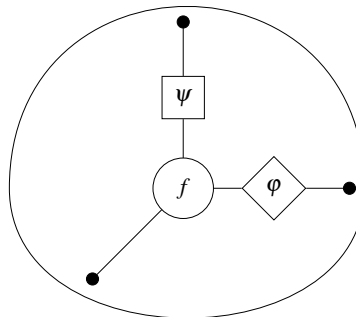


Figure 2: A composite factor  $f'$ .

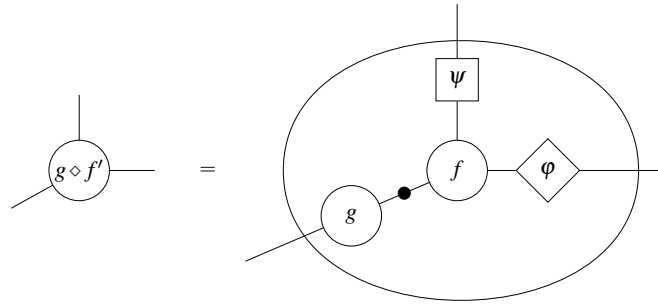
Figure 2 shows a simple example of a composite factor  $f'$ , constituted by another factor  $f$  with three exposed variables, two of which are composed with transformations, one deterministic ( $\varphi$ ) and one not ( $\psi$ ). Thinking statistically, we can interpret exposed variables (dangling wires) as representing *observed random variables*. We can draw a bubble around the whole of Figure 2 to obtain a new factor with three unobserved variables; nonetheless these variables are *observable*, which we depict by terminating them with a black dot. As a horizontal 1-cell in  $\mathbb{F}\mathbb{G}$ , the resulting factor has the type  $f_0 : 0 \rightarrow 0$ , where 0 denotes the empty set.



The drawing of the bubble represents a 2-cell  $f_0 \Rightarrow f'$ . We can think of this 2-cell, being directed from  $f_0$  to  $f'$ , as "reaching into  $f_0$ ", pulling observable wires out to expose them as the wires of  $f'$ .

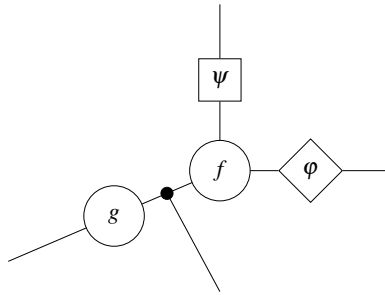
We can compose  $f$  with some factor  $g$  (with a compatible interface), repurposing the black dot — or

*spider*, in the terminology of Coecke and Kissinger [18] — to represent the locus of composition:

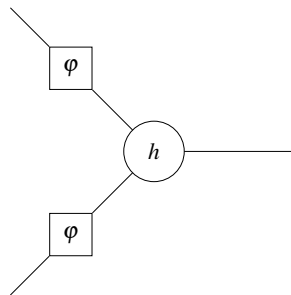


(In this way, we can think of a one-legged spider, representing an observable variable, as a locus of *potential* composition.)

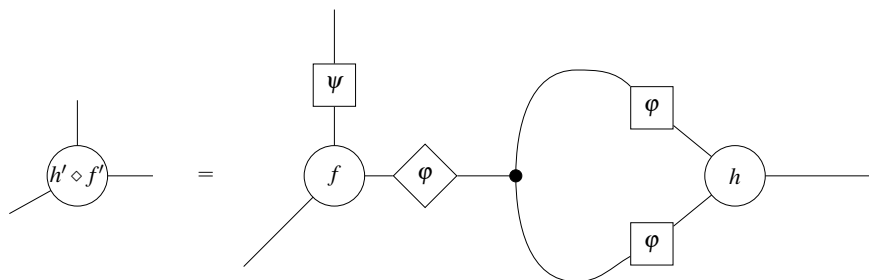
We can reach inside  $g \diamond f'$  to expose the internal wire:



And this gives us another way to understand the deterministic morphisms: as those that can slide past spiders, multiplying on the way into factors. For example, suppose we have another composite factor  $h'$  involving  $\varphi$ , such as this:

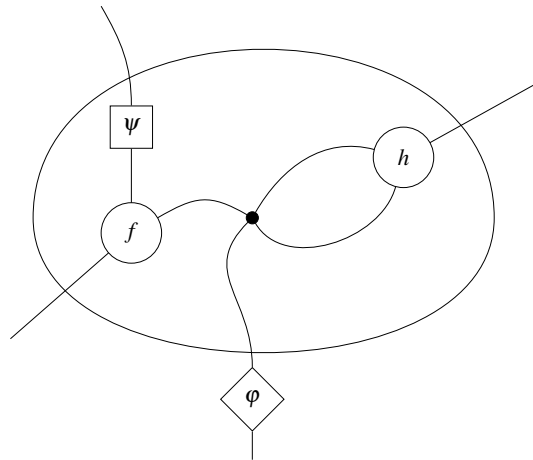


Then we could compose along  $\varphi$  with  $f'$  to obtain  $h' \diamond f'$ :



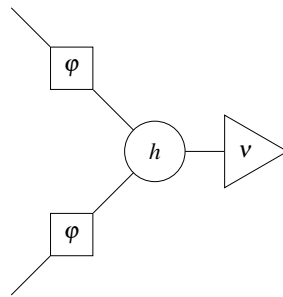
Notice that here we have made use of the ability (but not necessity!) of copy-composition to couple multiple variables together.

If again we expose the composed variable, we can pull  $\varphi$  outside, too:



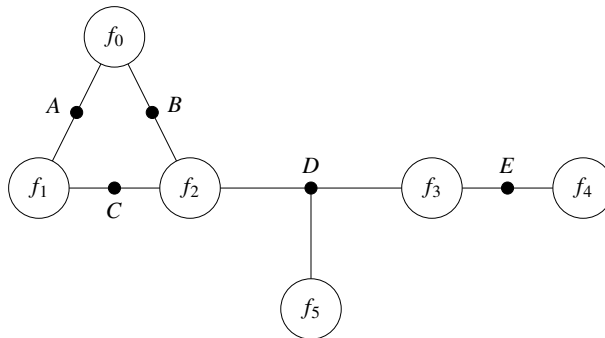
Notice that, by using spiders, there would be no ambiguity if we omitted the outer bubble: observed variables are those that are not terminated in a one-legged spider; and deterministic transformations “push forward” from exposed variables to multiply through spiders in the direction of their target factors.

We can use non-deterministic morphisms to render observable variables *unobservable*, by ‘marginalizing’ them out. For instance, here we have marginalized out one of the observed variables of  $h'$  by transforming it with a state  $v$ :



Here, we have used the standard string-diagrammatic representation of a state in  $\mathcal{C}$ .

Finally, let us depict the factor graph of Figure 1 in this language:



Note how similar this depiction is to the non-compositional form of Figure 1: the major change is the replacement of ‘variable’ nodes with (labelled) two-legged spiders, making the distinction between

observed and observable variables. Beyond this distinction, the syntax of  $\mathbb{FG}$  tells us precisely what kinds of compositions are allowed — including supplying notions of *transformation* of factors.