Reprinted from

HANDBOOK OF ALGEBRA

Volume 1

edited by
M. HAZEWINKEL
CWI, Amsterdam

1996

ELSEVIER
AMSTERDAM · LAUSANNE · NEW YORK · OXFORD · SHANNON · TOKYO

Categorical Structures

Ross Street

School of Mathematics, Physics, Computing and Electronics, Macquarie University, New South Wales 2109, Australia
e-mail: street@macadam.mpce.mq.edu.au

Contents

HANDBOOK OF ALGEBRA, VOL. 1
Edited by M. Hazewinkel

Introduction

Category theory is a young subject yet has, by now, contributed its share of substantial theorems to the vast body of mathematics. In certain areas, I consider that it has also managed to revolutionize thinking. Examples of such areas, and the innovative categorical concepts, are:

- *homological algebra:* abelian category [F, Sch, Gt];
- *universal algebra:* triple (= monad), sketch [ML2, Sch, BW];
- *algebraic geometry:* scheme, topos [SGA, Sch, Gd, Jt, MLM];
- *set theory:* elementary topos [Jt, BW, MLM];
- *enumerative combinatorics:* Joyal species [Joy].

These matters are well covered by the indicated accessible literature; therefore, it is not the purpose of this article to repeat them. I shall be concerned more with categories as vital mathematical structures (as emphasized by Ehresmann [Eh1, Eh2] and Lawvere [L]), rather than with traditional category *theory*.

In topology texts, we read that the spaces were designed to carry continuity to a useful conceptual level. Yet, categories are *two* steps away from naturality, the concept they were designed to formalize. The intermediate notion, functor, is the expected kind of morphism between categories. From the very study of the established practice of routinely specifying morphisms along with each mathematical structure, we were presented, in the 1940's, with an extra dimension: morphisms between morphisms. We were naturally led by naturality to objects, arrow *and* 2-cells. Topology had its analogue: homotopies.

The reader will be assumed to have familiarity with categories, functors and natural transformations. My starting point is the introduction of 2-cells. I consider a category further equipped with 2-cells, but with no compositions apart from the composition of arrows already existing in the category; this is called a *derivation scheme*. With such a simple structure, this paper explores some fundamental interconnections involving:

- rewrite systems;
- free higher-order categories;
- cubes and simplexes;
- string diagrams, Penrose tensor notation, and braids;
- the d-simplex equations arising in the study of exactly soluble models in statistical mechanics and quantum field theory;
- homotopy theory;
- coherence in category theory.

Convention. The composite of arrows $\alpha\colon a \to b$, $\beta\colon b \to c$ in a category A will be written in the algebraic order $\alpha \circ \beta\colon a \to c$. The other order may be regarded as "evaluation", so that parentheses $\beta(\alpha)\colon a \to c$ will be used.

1. Graphs, and 2-graphs

Recall that a *(directed) graph* G consists of two sets G_0, G_1 and an ordered pair of functions $s, t\colon G_1 \to G_0$. Elements of G_0 are called *objects, vertices, or* 0*-cells*. Elements of G_1 are called *arrows, edges, or* 1*-cells*. Call $s(\gamma)$ the *source* of the arrow γ, call $t(\gamma)$ its *target* and denote this by $\gamma\colon\ s(\gamma) \to t(\gamma)$. For objects a, b of G, we write $G(a,b)$ for the set of arrows $\gamma\colon\ a \to b$. There is a category **Grph** whose objects are graphs; the arrows $f\colon\ G \to H$, called *graph morphisms*, are pairs of functions $f_0\colon\ G_0 \to H_0$, $f_1\colon\ G_1 \to H_1$ such that, if $\gamma\colon\ a \to b$ in G, then $f_1(\gamma)\colon\ f_0(a) \to f_0(b)$ in H.

The *opposite* of a graph G is the graph G^{op} obtained from G by interchanging the functions s, t.

Each category A has an underlying graph (since a category has a set A_0 of objects and a set A_1 of arrows) which we also denote by A. The free category on (or generated by) a graph G is *the category* **F**G *of paths in* G, described as follows. The objects of $\mathbf{F}G$ are the objects of G. A *path* from a_0 to a_n of length $n \geqslant 0$ is a $(2n+1)$-plet $(a_0, \gamma_1, a_1, \gamma_2, \dots, \gamma_n, a_n)$:

$$a_0 \xrightarrow{\gamma_1} a_1 \xrightarrow{\gamma_2} a_2 \xrightarrow{\gamma_3} \cdots \xrightarrow{\gamma_n} a_n$$

where $\gamma_m\colon\ a_{m-1} \to a_m$ in G for $0 < m \leqslant n$. An arrow $\alpha\colon\ a \to b$ in $\mathbf{F}G$ is a path from a to b of any length $\ell(\alpha) \geqslant 0$. Composition of paths is given by

$$\begin{aligned}&(a_0, \gamma_1, a_1, \dots, \gamma_n, a_n) \circ (b_0, \delta_1, b_1, \dots, \delta_n, b_n)\\&\quad = (a_0, \gamma_1, a_1, \dots, \gamma_n, a_n, \delta_1, b_1, \dots, \delta_n, b_n)\end{aligned}$$

for $a_n = b_0$. So $\ell(\alpha \circ \beta) = \ell(\alpha) + \ell(\beta)$. It is convenient to identify the edge $\gamma\colon\ a \to b$ of G with the path $(a, \gamma, b)\colon\ a \to b$, and to denote the path $(a)\colon\ a \to a$ of length 0 hy $1_a\colon\ a \to a$ (as we do for identity arrows in any category). For $n > 0$, we then have

$$(a_0, \gamma_1, a_1, \gamma_2, \dots, \gamma_n, a_n) = \gamma_1 \circ \gamma_2 \circ \cdots \circ \gamma_n.$$

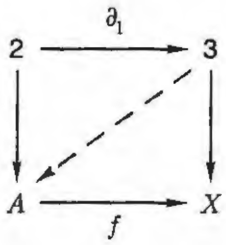
A category is called *free* when it is isomorphic to a category $\mathbf{F}G$ of paths in some graph G. For example, the category $\mathbf{N}$ which has one object 0, natural numbers $n\colon\ 0 \to 0$ as arrows, and addition as composition, is free. Each free category A has a length functor

$$\ell\colon\ A \to \mathbf{N};$$

the generating graph has the same objects as A, but only the arrows of length 1. The generating graph for $\mathbf{N}$ is a terminal object in the category **Grph**.

Let **2** denote the free category on the graph with two objects 0, 1, and one arrow $0 \to 1$. Let **3** denote the free category on the graph with three objects 0, 1, 2, and two arrows $0 \to 1 \to 2$. Let $\partial_i\colon \mathbf{2} \to \mathbf{3}$, $i = 0, 1, 2$, denote the functor which is injective on objects and does not have $i = 0, 1, 2$ in the image.

A functor $f\colon\ A \to X$ is said to be *ulf* (for "unique lifting of factorizations") when each commutative square of functors



has a unique filler, as shown by the dashed functor, making the two triangles commute. A category A is free if and only if there exists an ulf functor $\ell\colon A \to \mathbf{N}$.

For any category A, there is a functor **comp**: $\mathbf{F}A \to A$ given hy "composing the paths":

$$\mathbf{comp}(\xi) = \gamma_1 \circ \cdots \circ \gamma_n \quad \text{for } \xi = (a_0, \gamma_1, a_1, \ldots, \gamma_n, a_n).$$

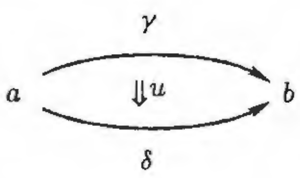
In fact, the category structure on the graph A is encapsulated by the graph morphism **comp**: $\mathbf{F}A \to A$; the precise statement is that the underlying functor from the category **Cat** of categories to **Grph** is monadic (or "tripleable").

The *chaotic graph* X_{ch} on a set X has source and target given by the first and second projections $X \times X \to X$. There is a unique category structure on X_{ch} so it is also called the *chaotic category* on X. The *discrete graph* X_{d} on the set X has source and target both given by the unique function $\varnothing \to X$. The *discrete category* on X is the free category $\mathbf{F}X_{\text{d}}$ on X_{d}; its source and target are hoth the identity function $1_X\colon X \to X$ of X.

Let $\pi_0 G$ denote the set of *connected components* of G; it is obtained from G_0 by identifying objects which have an arrow hetween them. Clearly $\pi_0 G = \pi_0 \mathbf{F}G$.

A *2-graph* G consists of three sets G_0, G_1, G_2 and four functions $s, t\colon G_1 \to G_0$, $s_1, t_1\colon G_2 \to G_1$ such that $s_1 \circ s = t_1 \circ s$ and $s_2 \circ t = t_1 \circ t$. The last two functions are denoted by $s, t\colon G_2 \to G_0$. Terminology for the graph $s, t\colon G_1 \to G_0$ is used for the 2-graph. Also, the elements u of G_2 are called *2-cells*; when $\gamma, \delta\colon a \to b$ and $s_1(u) = \gamma$, $t_1(u) = \delta$, we write either

$u\colon \gamma \Rightarrow \delta\colon a \longrightarrow b$ or



Write $G(a, b)$ for the graph whose objects are arrows $\gamma\colon a \to b$, and whose arrows are 2-cells $u\colon \gamma \Rightarrow \delta\colon a \to b$. The graph $s_1, t_1\colon G_2 \to G_1$ is the disjoint union of the graphs $G(a, b)$, $a, b \in G_0$. There is a category **2-Grph** of 2-graphs whose arrows $f\colon G \to H$, called *2-graph morphisms*, are triplets of functions $f_i\colon G_i \to H_i$, $i = 0, 1, 2$, such that (f_0, f_1), (f_1, f_2) are graph morphisms.

The *opposite* G^{op} of a 2-graph is obtained by interchanging s, t: $G_1 \to G_0$. The *conjugate* G^{co} of G is obtained by interchanging s_1, t_1: $G_2 \to G_1$. There is also G^{coop}.

2. Derivation schemes, sesquicategories, and 2-categories

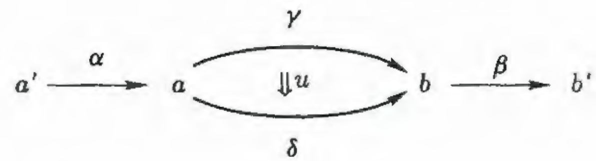
This section reviews concepts, selected from [S2] and [ES], which underpin 2-dimensional categories.

A *derivation scheme* D consists of a 2-graph D together with a category $\langle D\rangle$ whose underlying graph is s, t: $D_1 \to D_0$. We shall often provide the data for a derivation scheme D in a diagram

$$s_1, t_1\colon\ M \to A$$

where A is the category $\langle D\rangle$ and M is the set D_2. There is a category **DS** of derivation schemes whose arrows f: $D \to E$, called *derivation scheme morphisms*, are 2-graph morphisms for which (f_0, f_1): $\langle D\rangle \to \langle E\rangle$ is a functor.

Each 2-cell u: $\gamma \Rightarrow \delta$: $a \to b$ in a derivation scheme D can be thought of as a *rewrite rule* which labels the directed replacement of γ by δ. An *application* of the rule u is the replacement of any arrow of the form $\alpha \circ \gamma \circ \beta$ by $\alpha \circ \delta \circ \beta$. We label this application by the symbol $\alpha u \beta$: $\alpha \circ \gamma \circ \beta \Rightarrow \alpha \circ \delta \circ \beta$, and call it the *whiskering* of u by α, β as suggested by the following diagram.



It is harmless to identify u with its whiskering by identities. This gives a derivation scheme $\mathbf{w}D$ with tbe same category $\langle D\rangle$ and with the whiskered 2-cells; so $\mathbf{w}D$ contains D. A *derivation* in D is a finite sequence of applications of rules; more precisely, it is a path in the graph s_1, t_1: $(\mathbf{w}D)_2 \to D_1$. We obtain another derivation scheme $\mathbf{d}D$ with the same category $\langle D\rangle$ and with derivations as 2-cells. We write $(\mathbf{d}D)(a, b)$ for the path category of the graph $(\mathbf{w}D)(a, b)$. In fact, $\mathbf{d}D$ is more richly structured than a mere derivation scheme, it is an example of a "sesquicategory".

A *sesquicategory* S consists of a derivation scheme S and a functor

$$S(-,-)\colon\ \langle S\rangle^{\mathrm{op}} \times \langle S\rangle \to \mathbf{Cat}$$

whose composite with the functor obj: $\mathbf{Cat} \to \mathbf{Set}$ is the homfunctor of the category $\langle S\rangle$, and whose value at an object $(a, b) \in \langle S\rangle^{\mathrm{op}} \times \langle S\rangle$ is a category with underlying graph $S(a, b)$. We now write $S(a, b)$ for the category and not just the graph; the composition of $S(a, b)$ is called *vertical composition* and denoted by $\bullet$. For each pair of arrows α: $a' \to$

a, β: $b \to b'$, a functor $S(\alpha,\beta)$: $S(a,b) \to S(a',b')$ has its value at u: $\gamma \Rightarrow \delta$: $a \to b$ denoted by

$$\alpha \circ u \circ \beta\colon\ \alpha \circ \gamma \circ \beta \Rightarrow \alpha \circ \delta \circ \beta\colon\ a' \to b'$$

where $\circ$ between 1-cells is composition in the category $\langle S \rangle$. Let $\langle\langle S \rangle\rangle$ denote the category whose underlying graph is s_1, t_1: $S_2 \to S_1$ and whose composition is vertical composition $\bullet$.

There is a category **Sqc** of sesquicategories; the arrows, called *sesquifunctors*, are 2-graph morphisms which preserve all the compositions and identities.

Each sesquicategory S gives rise to a category $\mathbf{q}S$, called the *quotient category* of S. The objects are the objects of S. The set of arrows is the set of components of the category $\langle\langle S \rangle\rangle$. Composition is induced by that of $\langle S \rangle$ (this uses the compatibility of $\langle S \rangle$ composition with existence of 2-cells).

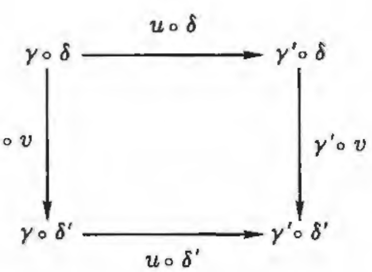
A *2-category* K [Eh1, Eh2] is a sesquicategory K such that, for all u: $\gamma \Rightarrow \gamma'$: $a \to b$, v: $\delta \Rightarrow \delta'$: $b \to c$, the following equation holds:

$$(u \circ \delta) \bullet (\gamma' \circ v) = (\gamma \circ v) \bullet (u \circ \delta').$$

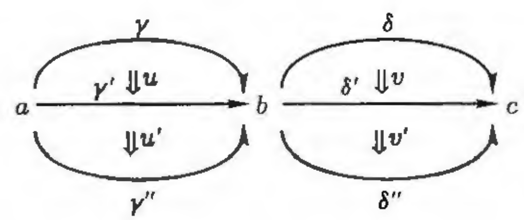
The 2-cell given by either side of the last equation is denoted by

$$u \circ v\colon\ \gamma \circ \delta \Rightarrow \gamma' \circ \delta'\colon\ a \to c$$

(and called the *horizontal composite* of the 2-cells u, v).
(HC)



It follows that the *middle-four-interchange law* holds: that is, for each diagram



in K, there is an equality

$$(u \bullet u') \circ (v \bullet v') = (u \circ v) \bullet (u' \circ v').$$

So, horizontal composition $- \circ -\colon K(a,b) \times K(b,c) \to K(a,c)$ is a functor. There is a category **2-Cat** of 2-categories; the arrows, now called *2-functors*, are sesquifunctors.

The basic example of a 2-category is **Cat**: its objects are categories (subject to some size restriction, if the reader feels this is needed), arrows are functors, and 2-cells are natural transformations [Gt], Appendix. Just as one considers additive categories, which are categories whose homsets are *enriched* in the monoidal category of abelian groups, we can describe 2-categories as categories whose homsets are enriched in **Cat** (with cartesian product as tensor product); see [EK] for precise definitions. Some connection between 2-categories and homotopy theory can be found in [GZ]. The connection between 2-categories and derivations in rewrite systems was made in [Bns].

Each sesquicategory S yields a 2-category $\mathbf{f}S$ by forcing commutativity in the squares (HC). This can be described by constructing a new derivation scheme E which will provide rewrite rules for arrows in S. Take $\langle E \rangle = \langle\langle S \rangle\rangle$. Take E_2 to be the subset of $S_2 \times S_2$ consisting of those pairs (u, v) of nonidentity 2-cells with $t(u) = s(v)$, and where $s_1, t_1\colon E_2 \to \langle S \rangle_1$ take (u, v) to the lower, upper paths around the above square (HC).

$$(u, v)\frac{(\gamma \circ v) \bullet (u \circ \delta')}{(u \circ \delta) \bullet (\gamma' \circ v)}.$$

Then form the quotient category $\mathbf{qd}E$ of the sesquicategory $\mathbf{d}E$. The objects of $\mathbf{qd}E$ are the arrows of S. Our 2-category $\mathbf{f}S$ is given by $\langle \mathbf{f}S \rangle = \langle S \rangle$ and $\langle\langle \mathbf{f}S \rangle\rangle = \mathbf{qd}E$. There is a canonical sesquifunctor $S \to \mathbf{f}S$, and composition with it establishes a bijection between 2-functors $\mathbf{f}S \to K$ and sesquifunctors $S \to K$, for all 2-categories K.

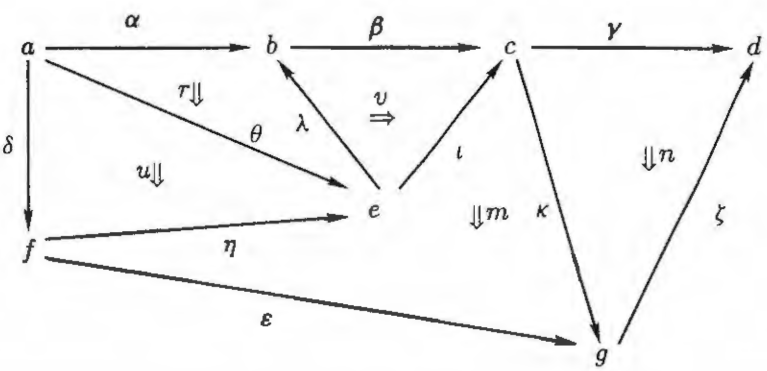
For any derivation scheme D, we can apply the construction of the last paragraph to the sesquicategory $S = \mathbf{d}D$ where the 2-cells are derivations in D and so have length. It is possible then to replace the derivation scheme E by the sub-derivation-scheme $\uparrow D$ of E whose 2-cells (u, v) are restricted to those with u, v both derivations of length 1. We call $\uparrow D$ the *lift* of D.

For any derivation scheme D, we obtain a 2-category $\mathbf{fd}D$. Two derivations in D are called *equivalent* when they are identified by the canonical sesquifunctor $\mathbf{d}D \to \mathbf{fd}D$; this means there is an undirected sequence of applications of the rules of $\uparrow D$ taking one derivation to the other.

3. Pasting, computads, and free 2-categories

Repeated horizontal and vertical composition in a 2-category K determine a more general operation called *pasting*. For example, consider the following diagram in K.

(P)



A 2-cell in a region means that its source and target are given by the composites of the indicated paths: for example, we have $v : \lambda \circ \beta \Rightarrow \iota$, $m : \eta \circ \iota \circ \kappa \Rightarrow \varepsilon$, and $r : \alpha \Rightarrow \theta \circ \lambda$. (Care is needed in placing the double arrow in each region so that it is clear which path is intended to be the source and which the target. If the arrow for r had pointed from left to right instead of downward, the result would be meaningless.) The 2-cells of the diagram (P) can be whiskered in such a way as to obtain a path from $\alpha \circ \beta \circ \gamma$ to $\delta \circ \varepsilon \circ \zeta$ of length 5 in the underlying graph of the category $K(a, d)$; for example,

$$\alpha \circ \beta \circ \gamma \xrightarrow{r \circ \beta \circ \gamma} \theta \circ \lambda \circ \beta \circ \gamma \xrightarrow{\theta \circ v \circ \gamma} \theta \circ \iota \circ \gamma \xrightarrow{u \circ \iota \circ \gamma}$$
$$\delta \circ \eta \circ \iota \circ \gamma \xrightarrow{\delta \circ \eta \circ \iota \circ n} \delta \circ \eta \circ \iota \circ \kappa \circ \zeta \xrightarrow{\delta \circ m \circ \zeta} \delta \circ \varepsilon \circ \zeta.$$

Another such path is

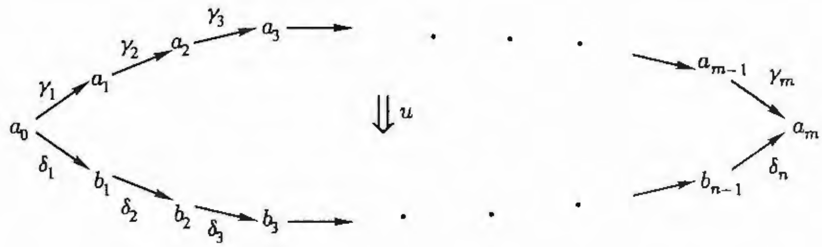
$$\alpha \circ \beta \circ \gamma \xrightarrow{\alpha \circ \beta \circ n} \alpha \circ \beta \circ \kappa \circ \zeta \xrightarrow{r \circ \beta \circ \kappa \circ \zeta} \theta \circ \lambda \circ \beta \circ \kappa \circ \zeta \xrightarrow{u \circ \lambda \circ \beta \circ \kappa \circ \zeta}$$
$$\delta \circ \eta \circ \lambda \circ \beta \circ \kappa \circ \zeta \xrightarrow{\delta \circ \eta \circ v \circ \kappa \circ \zeta} \delta \circ \eta \circ \iota \circ \kappa \circ \zeta \xrightarrow{\delta \circ m \circ \zeta} \delta \circ \varepsilon \circ \zeta.$$

We leave it as an exercise for the reader to check that these paths have the same composite in the category $K(a, d)$. Diagrams such as (P) are called *pasting diagrams*, and the 2-cell

$$(r \circ \beta \circ \gamma) \bullet (\theta \circ v \circ \gamma) \bullet (u \circ \iota \circ \gamma) \bullet (\delta \circ \eta \circ \iota \circ n) \bullet (\delta \circ m \circ \zeta) \colon \alpha \circ \beta \circ \gamma$$
$$\Rightarrow \delta \circ \varepsilon \circ \zeta \colon a \to d$$

is called the *pasting composite* of the diagram. Notice that, if we reversed the direction of the 2-cell r (say) in (P), we would no longer have a pasting diagram since no path in $K(a, d)$ could be made from it by whiskering the 2-cells.

A *computad* C consists of a graph $s, t \colon C_1 \to C_0$, denoted by $C^{\#}$, together with a derivation scheme $s_1, t_1 \colon C_2 \to \mathbf{F}C^{\#}$. The elements u of C_2 can be pictured as diagrams



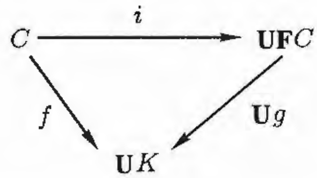
where the upper path is $s_1(u)$ and the lower is $t_1(u)$. A *computad morphism* $f\colon C \to C'$ is a triplet of functions $f_i\colon C_i \to C'_i$, $i = 0, 1, 2$, for which there is a morphism (f_0, f'_1, f_2) of the derivation schemes such that f'_1 agrees with f_1 on arrows of length 1. This gives a category **Cptd** of computads. Having given this precise definition, we can regard a computad as a derivation scheme C with $\langle C \rangle$ a free category, so long as we take care to remember that the computad morphisms preserve the length of 1-cells.

Each 2-category K has an underlying computad $C = \mathbf{U}K$ with $C^{\#}$ the underlying graph of the category $\langle K \rangle$, with

$$C_2 = \{(\xi, u, \eta) \mid \xi, \eta \text{ are paths in } C^{\#} \text{ and } u\colon \mathbf{comp}(\xi) \Rightarrow \mathbf{comp}(\eta) \text{ in } K\},$$

and with $s_1, t_1\colon C_2 \to \mathbf{F}\langle K \rangle$ taking (ξ, u, η) to ξ, η, respectively.

The *free* 2-*category* $\mathbf{F}C$ on the computad C is $\mathbf{fd}C$. There is an obvious inclusion computad morphism $i\colon C \to \mathbf{UF}C$. For each 2-category K and each computad morphism $f\colon C \to \mathbf{U}K$, there exists a unique 2-functor $g\colon \mathbf{F}C \to K$ such that the following triangle commutes.



This means that the functor $\mathbf{U}$ has a left adjoint $\mathbf{F}$. Taking $C = \mathbf{U}K$, we obtain a 2-functor **past**: $\mathbf{FU}K \to K$, called the *pasting operation* for the 2-category K. A 2-category structure on a computad C can be characterized in terms of an abstract pasting operation $\mathbf{UF}C \to C$. More precisely, the functor $\mathbf{U}$: **2-Cat** $\to$ **Cptd** is monadic.

This pasting operation will now be related to our previous discussion of the diagram (P). Suppose now that (P) is made up from data of a computad C. For example, there are 2-cells

$$v\colon (e, \lambda, b, \beta, c) \Rightarrow (e, \iota, c), \quad m\colon (f, \eta, e, \iota, c, \kappa, g) \Rightarrow (f, \varepsilon, g),$$

and

$$r\colon (a, \alpha, b) \Rightarrow (a, \theta, e, \lambda, b).$$

Whiskering the five 2-cells in the derivation scheme D of C, we obtain 2-cells

$$r \circ (b, \beta, c, \gamma, d), \quad (a, \theta, e) \circ v \circ (c, \gamma, d), \quad u \circ (e, \iota, c, \gamma, d),$$

$$(a, \delta, f, \eta, e, \iota, c) \circ n, \quad (a, \delta, f) \circ m \circ (g, \zeta, d)$$

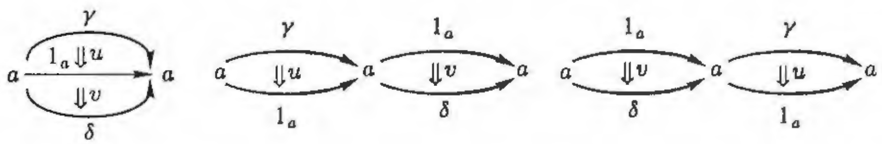
from $(a, \alpha, b, \beta, c, \gamma, d)$ to $(a, \delta, f, \varepsilon, g, \zeta, d)$ in $\mathbf{w}D$; they form a path in the graph $(\mathbf{w}D)(a, d)$. The connected component (with respect to (HC)) of this path gives a 2-cell

$$(a, \alpha, b, \beta, c, \gamma, d) \Rightarrow (a, \delta, f, \varepsilon, g, \zeta, d)\colon\ a \to d$$

in $\mathbf{F}C$. This is, of course, none other than the pasting composite of the diagram (P) in the 2-category $\mathbf{F}C$. Conversely, any other representative of this 2-cell in $\mathbf{F}C$ by a path in $(\mathbf{w}D)(a, d)$ leads us back to a planar diagram equivalent to (P). So a pasting diagram in C seems to provide a geometrically invariant way of depicting a 2-cell of $\mathbf{F}C$. For a 2-category K, the pasting operation **past**: $\mathbf{FU}K \to K$ assigns the pasting composite to the pasting diagram.

In general, however, when there are 2-cells which have source or target paths of length 0 in the computad C, the faithful geometric representation of 2-cells of $\mathbf{F}C$ by pasting diagrams breaks down. The reason is that the following three geometrically inequivalent pasting diagrams all represent the same 2-cell when

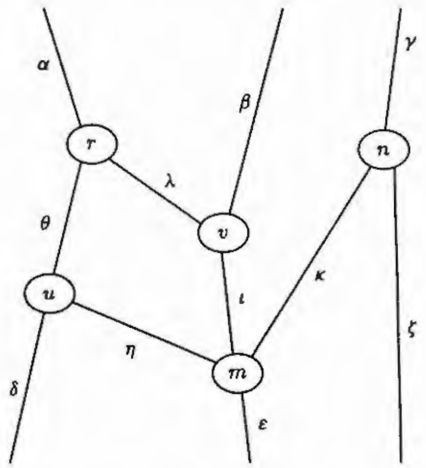
$$t_1(u) = s_1(v) = 1_a\colon\ a \to a.$$



We shall see below that this problem can be overcome by using the string diagrams which are planar dual to pasting diagrams.

4. Strings, and the terminal computad

Consider the planar dual of the pasting diagram (P) at the beginning of Section 3. Each 2-cell r, u, v, m, n becomes a node labeled by the same symbol; each arrow $\alpha, \beta, \ldots$ becomes an edge, called a *string*. A string is attached to a node when the original arrow formed part of the boundary of the region containing the 2-cell. Moreover, we require that the strings *progress* down the page from nodes that were source 2-cells towards nodes that were target 2-cells. The resultant graph, embedded in the plane, is called *a string diagram.*

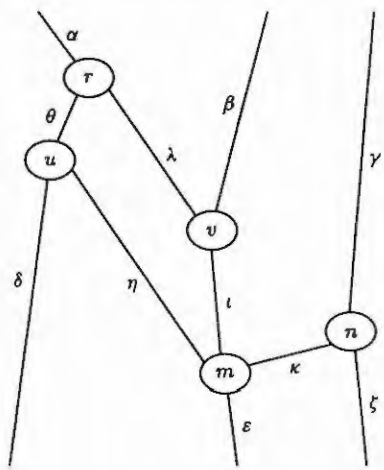


The *value* of this string diagram is the 2-cell $\alpha \circ \beta \circ \gamma \Rightarrow \delta \circ \varepsilon \circ \zeta$ obtained by breaking the diagram into horizontal layers with nodes at different levels in different layers. Reading from left to right, we obtain a horizontal composite of 2-cells from each layer; each node contributes its 2-cell, and each nodeless string contributes the identity 2-cell of its arrow. This gives the value of each layer. Then the values of the layers are composed vertically, reading down the page. For our example, we obtain:

$$(r \circ \beta \circ n) \bullet (\theta \circ v \circ \kappa \circ \zeta) \bullet (u \circ \iota \circ \kappa \circ \zeta) \bullet (\delta \circ m \circ \zeta).$$

The reader should enjoy checking that this agrees with the pasting composite of the pasting diagram (P) using the axioms for a 2-category.

The above string diagram can be deformed in the plane (as below) so as to preserve the strings' progression downward. The value remains the same [JS2].

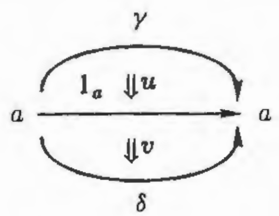


The value of this deformed string diagram is

$$(r \circ \beta \circ \gamma) \bullet (u \circ \lambda \circ \beta \circ \gamma) \bullet (\delta \circ \eta \circ v \circ \gamma) \bullet (\delta \circ \eta \circ \iota \circ n) \bullet (\delta \circ m \circ \zeta),$$

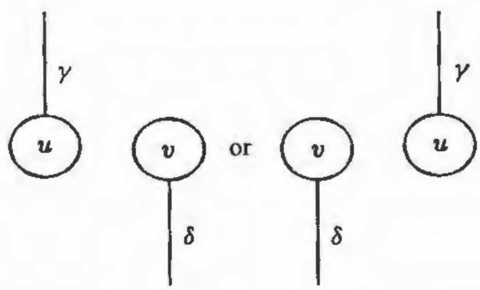
which is also equal to the pasting composite of (P).

Moreover, the string representation deals with the problem involving identities described at the end of Section 3. For suppose we have 2-cells u, v with $t_1(u) = s_1(v) = 1_a\colon a \to a$. Corresponding to the pasting diagram



we have the string diagram

whose value is $u \bullet v$ and which can be deformed to



which have the values $u \circ v$ and $v \circ u$, respectively. This suggests that the geometry of the string diagram provides a faithful representation of 2-cells in free 2-categories, which is indeed the case [JS2] as we shall explain in more detail below.

Just as it is of particular interest to consider the free category **N** on the terminal graph, it is also worth considering the free 2-category **M** on the terminal computad. Recall that **N** is a one-object category, and so is really just a monoid. Similarly, **M** is a one-object 2-category, and so is really just a strict monoidal category (that is, a monoid in the category **Cat** of categories and functors).

The terminal computad C_t is the terminal object in the category **Cptd**. The graph $C_t^\#$ is the terminal graph. So $\mathbf{F}C_t^\# = \mathbf{N}$. There must be exactly one 2-cell for each possible source and target path; so the derivation scheme of C_t is the chaotic graph on the set $\{0, 1, 2, \ldots\}$ of natural numbers. We write the 2-cells of $C_t^\#$ as m/n: $m \Rightarrow n$.

The derivation scheme $\mathbf{w}C_t$ has 2-cells of the form $(1, m/n, r)$: $1+m+r \Rightarrow 1+n+r$ obtained by whiskering m/n on the left by 1 and on the right by r. Thus the free 2-category **M** on C_t is obtained by taking paths of these 2-cells and identifying subject to condition (HC) of Section 2.

This gives the following direct description of **M** as a strict monoidal category. Consider the graph **W** whose vertices are natural numbers and whose edges $(l, m/n, r)$: $a \to b$ consist of natural numbers l, m, n, r with $l+m+r = a$ and $l+n+r = b$. Then consider the path category **FW**. We introduce the following "rewrite rule" on arrows of **FW** of length 2:

$$\frac{(l, m/n, r) \circ (l', m'/n', r')}{(l', m'/n', r' - n + m) \circ (l - m' + n', m/n, r)} \quad \text{for } l' + m' \leqslant l,$$

and, to exclude the case where the top and bottom are equal, we ask that not all of $l = l'$, $m = n = m' = n' = 0$ hold. This rule is a directed form of the condition (HC) as with the 2-cells of the lift $\uparrow C_t$. An application of this rewrite rule is the replacement of a path $\pi \circ \sigma \circ \pi'$ by $\pi \circ \tau \circ \pi'$ where σ is the top path and τ is the bottom path of the rule. To obtain **M**, identify arrows of **FW** when one arrow can be obtained from the other by a finite sequence of undirected applications of the rewrite rules. For objects c, d of **FW**, we have functors

$$c + -, - + d\colon \mathbf{FW} \to \mathbf{FW}$$

taking $(l, m/n, r)$: $a \to b$ to

$$(c + l, m/n, r)\colon c + a \to c + b,\ (l, m/n, r + d)\colon a + d \to b + d,$$

respectively. The identification of arrows in **FW** was introduced precisely so that these functors would induce partial functors for a functor

$$\mathbf{M} \times \mathbf{M} \to \mathbf{M}$$

which provides the tensor product for **M**; it is given on objects by addition of natural numbers.

We briefly consider the question of whether the (directed) rewrite rule above can be used to find "normal representatives" in **FW** for arrows in **M**. Notice that we do have "confluence" for the rewrite rules in the sense that, starting with a path in **W** of length 3 for which two rewrite rules can be applied, we can begin by applying either rule, yet continue applying rules to obtain a common result. For, suppose we have both $l' + m' \leqslant l$ and $l'' + m'' \leqslant l'$. Then $l'' + m'' \leqslant l' \leqslant l - m' \leqslant l - m' + n'$; so we have the following derivation.

$$\dfrac{\dfrac{(l,m/n,r)\circ(l',m'/n',r')\circ(l'',m''/n'',r)}{(l',m'/n',r'-n+m)\circ(l-m'+n',m/n,r)\circ(l'',m''/n'',r)}}{\dfrac{(l',m'/n',r'-n+m)\circ(l'',m''/n'',r''-n+m)\circ(l-m'+n'-m''+n'',m/n,r)}{(l'',m''/n'',r''-n+m-n'+m')\circ(l'-m''+n'',m'/n',r'-n+m)\circ(l-m'+n'-m''+n'',m/n,r)}}$$

Also, $(l' - m'' + n'') + m' = (l' + m') - m'' + n'' \leqslant l - m'' + n''$; so we have the following derivation.

$$\dfrac{\dfrac{(l,m/n,r)\circ(l',m'/n',r')\circ(l'',m''/n'',r)}{(l,m/n,r)\circ(l'',m''/n'',r''-n'+m')\circ(l'-m''+n'',m'/n',r')}}{\dfrac{(l'',m''/n'',r''-n'+m'-n+m)\circ(l-m''+n'',m/n,r)\circ(l'-m''+n'',m'/n',r)}{(l'',m''/n'',r''-n'+m'-n+m)\circ(l'-m''+n'',m'/n',r'-n+m)\circ(l-m''+n''-m'+n',m/n,r)}}$$

Notice that the derivations both lead to the same bottom line, yielding the desired confluence.

A path in **W** is called *reduced* when the rewrite rules cannot be applied to it. So a path $(l, m/n, r) \circ (l', m'/n', r')$ of length 2 is reduced when either $l < l' + m'$, or $m = n = m' = n' = 0$ and $l = l'$. An arbitrary path is reduced if and only if every path of length 2 through which it factors is reduced. Notice that, if $l' + m' \leqslant l$, then the path

$$(l', m'/n', r' - n + m) \circ (l - m' + n', m/n, r')$$

is reduced if $n' + m > 0$; so in this case, for paths of length 2, a reduced path is obtained in one application of a rewrite rule. For the case $n' + m = 0$, notice the derivation of length 2:

$$\dfrac{\dfrac{(m', 0/n, 0) \circ (0, m'/0, n)}{(0, m'/0, 0) \circ (0, 0/n, 0)}}{(0, 0/n, m') \circ (n, m'/0, 0)}.$$

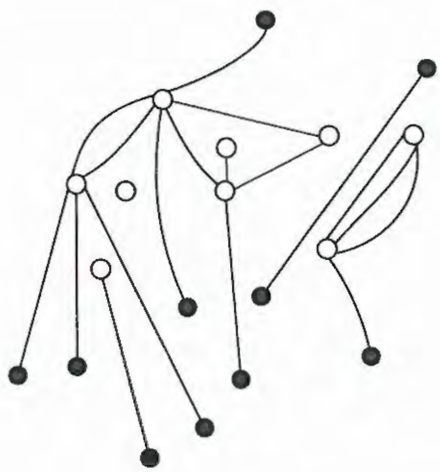
This is why we need the second sentence of the following result.

PROPOSITION 4.1 [ES]. *Let π: $a \to b$ be a path of length k in the graph* **W**. *Suppose π does not contain both an edge $(l, m/n, r)$ with $m = 0$ and an edge $(l', m'/n', r')$ with $n' = 0$. Then all derivations with source π, using the above rewrite rules, have length $\leqslant k(k-1)/2$. Moreover, π is equivalent to a unique reduced path.*

REMARK. Without the second sentence of the Proposition 4.1, the upper bound $k(k-1)/2$ must be increased (as shown by the above derivation of length 2 with $k = 2$). David Benson has advised me that $k(k-1)$ is an upper bound in the general case, and that this follows from his paper [Bns]. The complication is related to the one discussed at the end of Section 3, which reminds us to look at a string model for **M**.

A *plane graph* Γ is a compact topological subspace of $\mathbb{R}^2$ with a distinguished set Γ_0 of points whose complement $\Gamma - \Gamma_0$ in Γ is homeomorphic to a finite union of disjoint open intervals. The elements of Γ_0 are called *vertices* and the connected components of $\Gamma - \Gamma_0$ are called *edges*. We say that (x, y) is *above* (x', y') in $\mathbb{R}^2$ when $y' \leqslant y$; *below* means the reverse. The plane graph Γ is called *progressive* when aboveness is a total (linear) order on each edge. Progressive plane graphs are directed graphs: the source and target of an edge are the vertices in the closure of the edge; the source is above the target.

A *progressive plane graph with boundary* consists of a progressive plane graph Γ with a distinguished set $\mathbf{i}\Gamma$ of vertices such that each vertex in $\partial\Gamma = \Gamma_0 - \mathbf{i}\Gamma$ is in the closure of precisely one edge, and $\mathbf{i}\Gamma$ is an interval in the aboveness order on Γ_0 (that is, if p, q, r are vertices with p above q and q above r, then $p, r \in \mathbf{i}\Gamma$ implies $q \in \mathbf{i}\Gamma$). Notice tbat $\partial\Gamma$ is the disjoint union of the subset $\mathbf{s}\Gamma$ of those vertices which are sources and the subset $\mathbf{t}\Gamma$ of those vertices which are targets. For example, in the progressive plane graph depicted below, the white nodes provide an acceptable set $\mathbf{i}\Gamma$; so the black nodes constitute $\partial\Gamma$, the cardinality of $\mathbf{s}\Gamma$ is two, and the cardinality of $\mathbf{t}\Gamma$ is eight.



Of course, tbe size of the nodes is exaggerated for visibility. It is customary to omit the boundary (black) nodes from the picture, leaving loose the single edge having it in the closure.

Suppose Γ, Γ' are progressive plane graphs with boundary. We say that Γ is a *deformation* of Γ' when there exists a homeomorphism h: $\mathbb{R}^2 \to \mathbb{R}^2$ such that $h(\Gamma) = \Gamma'$, $h(\partial\Gamma) = \partial\Gamma'$, and h preserves the aboveness order on edges.

Now we give the geometric model of the strict monoidal category **M**. The objects are natural numbers. An arrow $[\Gamma]$: $m \to n$ is a deformation class of progressive plane graphs with boundary such that the cardinalities of $\mathbf{s}\Gamma$, $\mathbf{t}\Gamma$ are m, n, respectively. We define the composite $[\Gamma] \circ [\Lambda]$: $m \to p$ of arrows $[\Gamma]$: $m \to n$, $[\Lambda]$: $n \to p$ by choosing representatives Γ, Λ such that

$$\mathbf{s}\Gamma = \mathbf{t}\Lambda = \{(k, 0) \colon k = 1, 2, \ldots, n\},$$

with $\Gamma - \mathbf{t}\Gamma$ contained in the upper half plane and $\Delta - \mathbf{s}\Delta$ in the lower half plane; then put

$$[\Gamma] \circ [\Lambda] = [\Gamma \cup \Lambda]$$

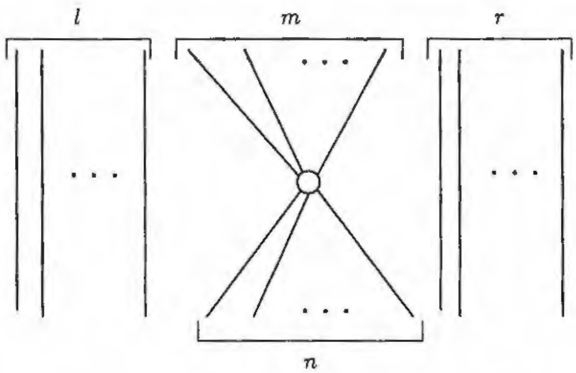
where $(\Gamma \cup \Lambda)_0 = (\Gamma_0 \cup \Lambda_0) - \mathbf{t}\Gamma$ and $\partial(\Gamma \cup \Lambda) = (\partial\Gamma \cup \partial\Lambda) - \mathbf{t}\Gamma$. We define the tensor product

$$[\Gamma] \otimes [\Gamma'] \colon m + m' \to n + n'$$

of arrows $[\Gamma]$: $m \to n$, $[\Gamma']$: $m' \to n'$ by choosing the representatives Γ, Γ' to be contained in the left, right half plane (respectively); then put

$$[\Gamma] \otimes [\Lambda] = [\Gamma \cup \Lambda]$$

where $(\Gamma \cup \Lambda)_0 = \Gamma_0 \cup \Lambda_0$ and $\partial(\Gamma \cup \Lambda) = \partial\Gamma \cup \partial\Lambda$. There is a graph morphism $\mathbf{W} \to \mathbf{M}$ which is the identity on objects and takes the edge $(l, m/n, r)$ in **W** to the deformation class of the following graph.



This graph morphism extends to a functor $\mathbf{FW} \to \mathbf{M}$ which is the universal functor out of **FW** identifying the rewrite rules for paths in **W** [JS2].

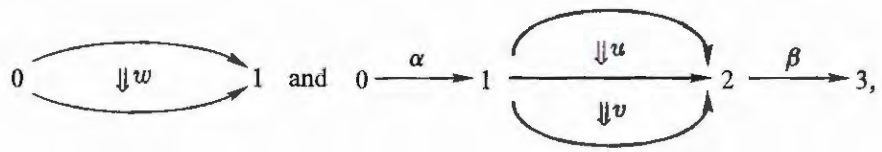
REMARK. When returning to the view of M as a 2-category, its single object will be denoted by 0, and horizontal, vertical composition will be denoted by $\circ, \bullet$ as usual in a 2-category, rather than by $\otimes, \circ$ with their usual meaning in a monoidal category.

5. Length 2-functors, and presentations of 2-categories

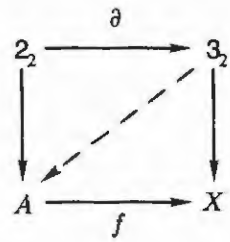
Each free 2-category A has a *length 2-functor* ℓ: $A \to \mathbf{M}$ induced by the unique computad morphism between the generating computads; recall that the generating computad of $\mathbf{M}$ is terminal. We now attempt to characterize free 2-categories in terms of the length 2-functor.

Each 2-functor ℓ: $A \to \mathbf{M}$ determines a computad $\ell^{-1}(C_t)$ which is the subcomputad of $\mathbf{U}A$ with the same objects, the arrows γ with $\ell(\gamma) = 1$, and the 2-cells u: $\alpha \Rightarrow \beta$ with $\ell(u)$ represented by the edge $(0, \ell(\alpha)/\ell(\beta), 0)$ of $\mathbf{W}$. If A is free and ℓ is its length 2-functor then A is free on the computad $\ell^{-1}(C_t)$.

Let $\mathbf{2}_2$ and $\mathbf{3}_2$ denote the free 2-categories on the computads depicted by



respectively, and let ∂: $\mathbf{2}_2 \to \mathbf{3}_2$ be the 2-functor which takes w to $\alpha \circ (u \bullet v) \circ \beta$. A 2-functor f: $A \to X$ is said to be *ulf* when each commutative square



can be uniquely filled by a 2-functor as indicated by the dashed arrow.

A computad (or derivation scheme) is called *tight* when there are no 2-cells u: $\alpha \Rightarrow \beta$: $a \to a$ with α the identity of a. (From the rewrite view of C, this is a mild requirement, since the possibility of rewriting nothing as something is seldom desirable as it leads to infinite derivations.) Let $\mathbf{M}'$ denote the free 2-category on the sub-computad of the computad C_t consisting of those 2-cells m/n: $m \Rightarrow n$ with $m \neq 0$.

PROPOSITION 5.1. *A 2-category A is free on a tight computad if and only if there exists an ulf 2-functor*

$$\ell\colon A \to \mathbf{M}'.$$

To characterize general free 2-categories, we take the string viewpoint. Let Γ denote a progressive plane graph with boundary, and let D be any computad. A *valuation* $\nu\colon \Gamma \to D$ of Γ in D consists of a pair of functions

$$\nu_0\colon \Gamma_1 \to D_1 \quad \text{and} \quad \nu_1\colon \mathbf{i}\Gamma \to D_2$$

such that, for each $x \in \mathbf{i}\Gamma$, one has

$$\nu_1(x)\colon \nu_0(e_1) \circ \nu_0(e_2) \circ \cdots \circ \nu_0(e_m) \Rightarrow \nu_0(f_1) \circ \nu_0(f_2) \circ \cdots \circ \nu_0(f_n)$$

where $e_1, \dots, e_m$ are the edges with target x ordered from left to right in the plane, and $f_1, \dots, f_m$ are the edges with source x also ordered from left to right. A *string diagram* in D is a pair (Γ, ν) consisting of a progressive plane graph Γ with boundary and a valuation $\nu\colon \Gamma \to D$. If (Γ, ν) is a string diagram in D and Γ' is a deformation of Γ then there is an obvious way to obtain a valuation ν' on Γ'; in this case, (Γ', ν') is called a *deformation of the string diagram* (Γ, ν). Write $[\Gamma, \nu]$ for the deformation class of (Γ, ν).

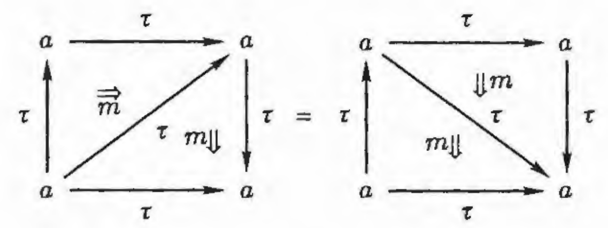
PROPOSITION 5.2. *A 2-category A is free on some computad if and only if there exists a 2-functor $\ell\colon A \to \mathbf{M}$ such that, for each string diagram (Γ, ν) in the computad $\ell^{-1}(C_t)$, there exists a unique 2-cell u in A with $\ell(u) = [\Gamma, \nu]$.*

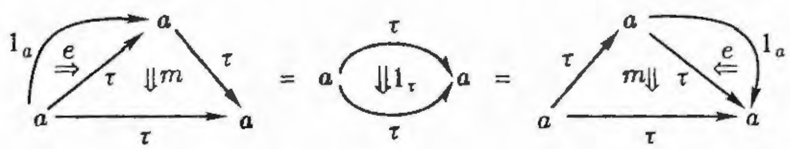
Suppose A in any 2-category. By a *valuation* $\nu\colon \Gamma \to A$ and a *string diagram* in A, we mean a valuation and a string diagram in the computad $\mathbf{U}A$. Suppose (Γ, ν) is any string diagram in A. By Proposition 5.2, there exists a unique 2-cell u in $\mathbf{FU}A$ with $\ell(u) = [\Gamma, \nu]$. The *value* $\nu(\Gamma)$ of the string diagram (Γ, ν) in A is the value of u under the 2-functor **past**: $\mathbf{FU}A \to A$; that is,

$$\nu(\Gamma) = \mathbf{past}(u).$$

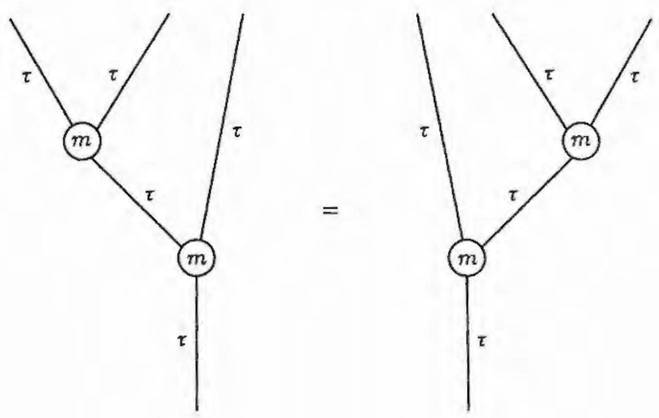
Now that we have some understanding of free 2-categories, we can contemplate presentations of 2-categories. A *presentation* of a 2-category consists of a computad C and a relation R on the set $(\mathbf{F}C)_2$ of 2-cells of the free 2-category $\mathbf{F}C$. (The elements (α, β) of R are often written as equations $\alpha = \beta$.) One obtains a 2-category A (unique up to isomorphism) by constructing the universal 2-functor $\mathbf{F}C \to A$ which identifies R-related 2-cells; then (C, R) is called a *presentation of* A. Of course, C can be identified with a subcomputad of $\mathbf{U}A$.

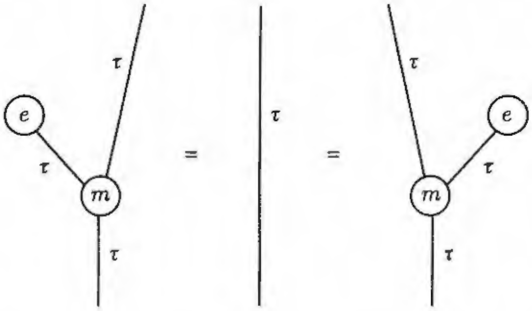
EXAMPLE 1. *Monads*. Consider the computad C with one object a, one arrow $\tau\colon a \to a$, and two 2-cells $e\colon 1_a \Rightarrow \tau$, $m\colon \tau \circ \tau \Rightarrow \tau$. While this computad is not tight, its conjugate C^{co} is tight, so all views of $\mathbf{F}C$ are available. Consider the relations R given as follows.





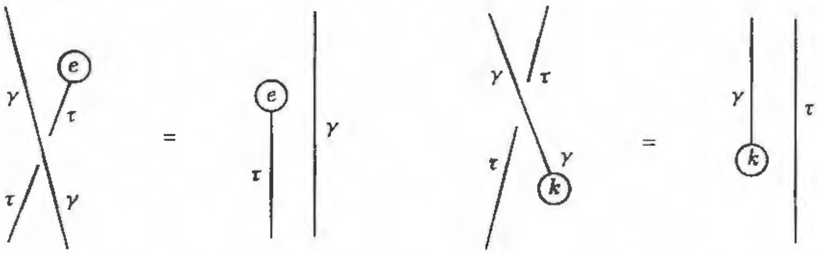
These relations can also be drawn using string diagrams, as follows.

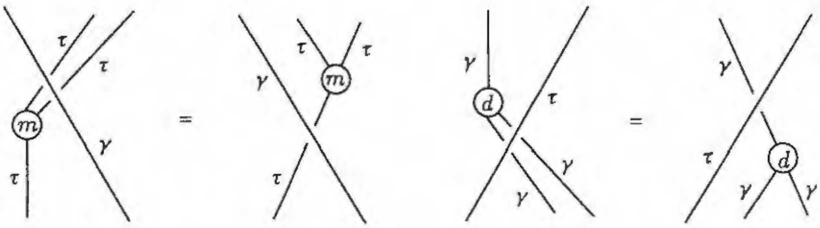




Let **Mnd** denote the 2-category with one object a, and with homcategory $\mathbf{Mnd}(a, a)$ equal to the category of finite ordinals and order-preserving functions; horizontal composition is ordinal sum. The (C, R) provides a presentation for the 2-category **Mnd** via the interpretation of τ as the ordinal 1, and e: $0 \to 1$, m: $2 \to 1$ as the unique functions. To give a 2-functor $\mathbf{Mnd} \to K$ into a 2-category K is to give a *monad* in K.

EXAMPLE 2. *Distributive laws between monads and comonads.* As a natural example of a computad C which is neither tight nor has a tight conjugate, we take one object a, two arrows τ, γ: $a \to a$, and five 2-cells e: $1_a \Rightarrow \tau$, m: $\tau \circ \tau \Rightarrow \tau$, k: $\gamma \Rightarrow 1_a$, d: $\gamma \Rightarrow \gamma \circ \gamma$, r: $\gamma \circ \tau \Rightarrow \tau \circ \gamma$. It will make the relations we are about to consider look more geometrically appealing if, in the string diagrams, we depict the 2-cell r as a cross-over of string γ over string τ, rather than as a node. Let R consist of the relations for e: $1_a \Rightarrow \tau$, m: $\tau \circ \tau \Rightarrow \tau$ as in Example 1, the relations given by inverting the string diagrams for e, m and replacing e, m by k, d, and the following four extra relations.



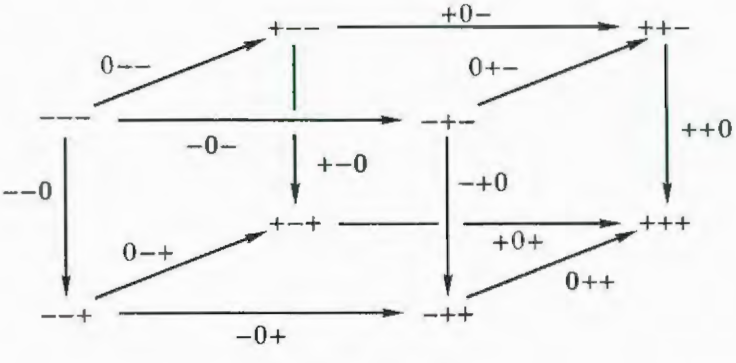


For a 2-category A, the computad morphisms $C \to \mathbf{U}A$, which identify R-related 2-cells, are in bijection with objects a of A equipped with a monad τ, comonad γ, and a distributive law r between them [Bc, S1, BW].

6. Cubes, and Gray's tensor product of 2-categories

By way of application of the above ideas, we now consider structures arising from consideration of cubes of all dimensions. What could be more basic than rewriting a single given symbol, say "minus", by another, say "plus"? We begin with a computad which, in a sense, is a combinatorial version of the *interval*, so we denote it by $\mathbf{I}$. The graph $\mathbf{I}^{\#}$ has one vertex (which shall remain nameless), and two edges denoted by $-$ and $+$. Paths in this graph are words α in the symbols $-$ and $+$; such words of length n are in bijection with the 2^n vertices of the n-cube. There is only one 2-cell in $\mathbf{I}$ which we denote by $\mathbf{0} : - \Rightarrow +$.

An application of the rewrite rule $\mathbf{0} : - \Rightarrow +$ to a word α of length n can be identified with an edge of the n-cube; it is a word u of length n in the symbols $-, 0, +$ with precisely one 0 occurring. The position of the 0 in u is a position in α where there is a symbol $-$ and the target of u is obtained from α by changing this $-$ to a $+$. Derivations in the derivation scheme $\mathbf{I}$ are paths around the edges of the cube. So the 2-cells of the one object sesquicategory **dI** can be regarded as paths around some n-cube. Write $\mathbf{I}[n, 1]$ for the subderivation scheme of **dI** consisting of the words α in the symbols $-$, $+$ of length n.



The 3-cube $\mathbf{I}[3, 1]$.

For words $\alpha, \beta \in \mathbf{I}[n, 1]$, write $\alpha \leqslant \beta$ when α, β have the same length and α has the symbol $-$ in every position that β does. Clearly there exists a derivation $\alpha \to \beta$ if and only if $\alpha \leqslant \beta$. Moreover, any two derivations with the same source and target are equivalent. It follows that the homcategory of the free 2-category **fdI** on **dI** is a partially ordered set: it is a strict monoidal category whose tensor product is juxtaposition of words. If we take the full subcategory of this homcategory consisting of the words α of length n, we obtain a category $\mathbf{Cub}[n, 1]$, called the *n-cube with commutative 2-faces*.

However, we may not wish the 2-faces to commute. In other words, we may not wish to identify equivalent derivations. Let us examine the derivations in more detail. Suppose α, β are words of the same length n in the symbols $-, +$ and suppose $\alpha \leqslant \beta$. Write $\alpha \backslash \beta$ for the set of positions where α has $-$ and β has $+$. A derivation u of $\mathbf{I}$ from α to β can be identified with a listing $u = u_1 u_2 \dots u_k$ of the elements of $\alpha \backslash \beta$ (each application determines an element of $\alpha \backslash \beta$ which is the position of the symbol 0 and the order is that forced by composibility of the applications making up the derivation). With this notation it must be realized that the source and target of $u\colon\ \alpha \to \beta$ must be specified in order to fully determine the derivation. Put

$$\mathcal{V}(u) = \{(u_i, u_j)\colon\ i < j \text{ and } u_i < u_j\}.$$

Notice that, for derivations $u\colon\ \alpha \to \beta$, $v\colon\ \beta \to \gamma$, there is a partition of $\mathcal{V}(uv)$ as

$$\mathcal{V}(uv) = \mathcal{V}(u) + \{(u_i, v_j)\colon\ u_i < v_j\} + \mathcal{V}(v).$$

We shall now describe the lift derivation scheme $\uparrow \mathbf{I}$. The objects are words α in the symbols $-, +$. The arrows are derivations $u\colon\ \alpha \to \beta$ of $\mathbf{I}$. The 2-cells are oriented 2-faces of an n-cube which can be depicted in pure $-, 0, +$ notation as

$$\begin{array}{ccc}
\alpha - \beta - \gamma & \xrightarrow{\ \alpha\, 0\, \beta\, -\, \gamma\ } & \alpha + \beta - \gamma \\
{\scriptstyle \alpha - \beta\, 0\, \gamma}\downarrow & \overset{\alpha\, 0\, \beta\, 0\, \gamma}{\Rightarrow} & \downarrow{\scriptstyle \alpha + \beta\, 0\, \gamma} \\
\alpha - \beta + \gamma & \xrightarrow[\ \alpha\, 0\, \beta\, +\, \gamma\ ]{} & \alpha + \beta + \gamma
\end{array}$$

or in "position of 0" notation as

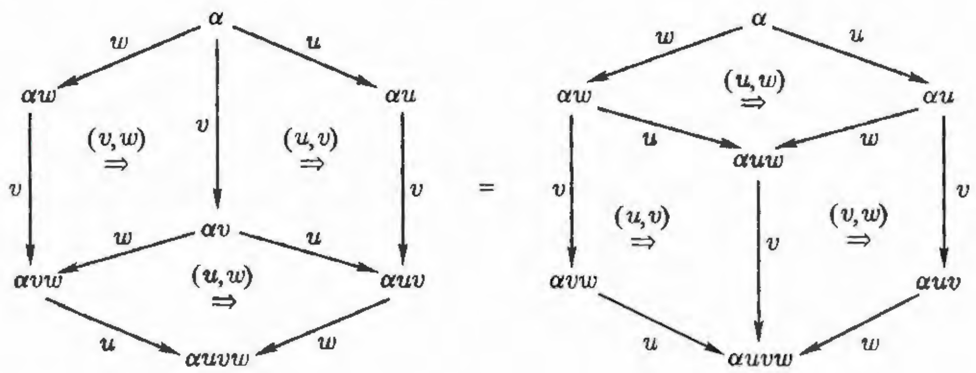
$$\begin{array}{ccc}
\alpha - \beta - \gamma & \xrightarrow{\ u\ } & \alpha + \beta - \gamma \\
{\scriptstyle v}\downarrow & \overset{(u,\ v)}{\Rightarrow} & \downarrow{\scriptstyle v} \\
\alpha - \beta + \gamma & \xrightarrow[\ u\ ]{} & \alpha + \beta + \gamma
\end{array}$$

where $u = \ell(\alpha) + 1,\ v = u + \ell(\beta) + 1$. Notice in the last square that

$$\mathcal{V}(vu) = \varnothing \subset \{(u,v)\} = \mathcal{V}(uv).$$

Write $\mathsf{I}[n,2]$ for the sub-derivation scheme of $\uparrow \mathsf{I}$ obtained by taking only the objects α of length n.

The *commuting* 3-*face relations* are the following relations on 2-cells in the free 2-category $\mathbf{F} \uparrow \mathsf{I}$: for each object α of $\uparrow \mathsf{I}$ with the symbol – in positions $u < v < w$,



where αu denotes the result of changing $-$ to $+$ in position u of α.

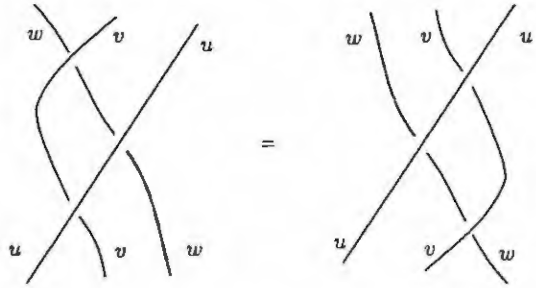
There is a 2-category $\mathbf{Cub}[n,2]$ defined as follows. The objects are words α of length n in the symbols $-, +$. For $\alpha \leqslant \beta$, the homcategory $\mathbf{Cub}[n,2](\alpha,\beta)$ is the ordered set of listings $u = u_1 u_2 \ldots u_k$ of the elements of $\alpha \backslash \beta$ where $u \leqslant u'$ if and only if $\mathcal{V}u \subseteq \mathcal{V}u'$. Otherwise, $\mathbf{Cub}[n,2](\alpha,\beta) = \varnothing$. Horizontal composition

$$\mathbf{Cub}[n,2](\alpha,\beta) \times \mathbf{Cub}[n,2](\beta,\gamma) \to \mathbf{Cub}[n,2](\alpha,\gamma)$$

is concatenation of listings which is order preserving (by the formula for $\mathcal{V}(uv)$).

PROPOSITION 6.1. *A presentation of the* 2-*category* $\mathbf{Cub}[n,2]$ *is provided by the computad* $\mathsf{I}[n,2]$ *subject to the commuting* 3-*face relations.*

Consequently, the 2-category $\mathbf{Cub}[n,2]$ is called the n-*cube with commutative* 3-*faces*. This 2-category was given in terms of generators and relations by Gray [Gy2] who used the positive part of the braid groups to show its homcategories were ordered (strong Bruhat order of the symmetric groups). To make a connection here with positive braids notice that the string diagrams for the commuting 3-face relations are as follows provided we depict the nodes as crossovers. (More will be said on this in later sections.)



The cube 2-categories arose in Gray's work in order to prove that his *tensor product of* 2*-categories* was a monoidal structure on the category **2-Cat**. (That is, that the tensor product is associative up to isomorphisms which satisfy certain axioms.) This tensor product

$$\otimes\colon\ \mathbf{2\text{-}Cat} \times \mathbf{2\text{-}Cat} \to \mathbf{2\text{-}Cat}$$

is not the product in the category **2-Cat**. One way to construct it is to first define it on the cube 2-categories by putting

$$\mathsf{Cub}[m,2] \otimes \mathsf{Cub}[n,2] = \mathsf{Cub}[m+n,2].$$

Then we need to observe:

PROPOSITION 6.2. *The full subcategory of* **2-Cat** *consisting of the* 2*-categories* $\mathsf{Cub}[n,2]$, $n = 0,1,2,\ldots$, *is dense. (In fact,* $\mathsf{Cub}[3,2]$ *alone suffices.)*

This means that every 2-category A is a canonical colimit

$$A \cong \mathrm{colim}_i \mathsf{Cub}[m_i, 2]$$

of cube 2-categories. Since we wish the functors $A \otimes -$, $- \otimes B$: **2-Cat** $\to$ **2-Cat** to preserve colimits, we are forced to the formula

$$A \otimes B \cong \mathrm{colim}_{i,j} \mathsf{Cub}[m_i + n_j, 2].$$

The fact that this approach leads to a biclosed monoidal structure on **2-Cat** follows from a general result of Day [Da2, Da3] on Kan extending tensor products along dense functors. Moreover, the **2-Cat**-valued homs, which provide right adjoints for $A \otimes -$ and $- \otimes B$, are easily described as "funny 2-functor 2-categories".

Before describing these, it is worth looking at the situation with the category **Cat**. Any category **V** with products becomes a monoidal category by taking the product as the tensor product; this is called the *cartesian monoidal structure* on **V**. Call **V** *cartesian closed* when, for all objects B, C of **V**, there exists an *exponential object* $[B, C]$

characterized up to isomorphism by the existence of a natural bijection between arrows $A \times B \to C$ and arrows $A \to [B, C]$. In the case $\mathbf{V} = \mathbf{Cat}$, of course, $[B, C]$ is the *functor category* whose objects are functors from A to B and whose arrows are natural transformations. Categories with homs enriched in the cartesian closed category **Cat** are precisely 2-categories.

However, for categories B, C, there is also the *funny functor category* $\{B, C\}$ whose objects are functors $f\colon B \to C$, and whose arrows $\theta\colon f \to g$ are families of arrows $\theta_b\colon f(b) \to g(b)$ in C indexed by the objects $b \in B$ (no naturality requirement!). There is a *funny tensor product* $A \otimes B$ of categories A, B such that functors $h\colon A \otimes B \to C$ are in natural bijection with functors $k\colon A \to \{B, C\}$. In fact, a category with homs enriched in the monoidal category **Cat** with the funny tensor product is more general than a 2-category; it is precisely a sesquicategory. (The funny tensor product was used recently [BG1, BG2] in studying Petri nets.)

The category **2-Cat** is cartesian closed. For 2-categories B, C, the exponential 2-category $[B, C]$ has 2-functors as objects, 2-natural transformations as arrows, and *modifications* as 2-cells. A category with homs enriched in **2-Cat**, with the cartesian structure, is called a *3-category*.

For 2-categories B, C, the *funny 2-functor 2-category* $\{B, C\}$ has 2-functors $f\colon B \to C$ as objects, transformations $\theta\colon f \to g$ as arrows, and modifications as 2-cells (this terminology will be discussed in Section 9 in the context of bicategories). There is a natural bijection between 2-functors $h\colon A \otimes B \to C$ (where $\otimes$ is Gray's tensor product of 2-categories) and 2-functors $k\colon A \to \{B, C\}$. So $\{B, -\}$ is a right adjoint for $- \otimes B$. A right adjoint for $A \otimes -$ is obtained using the canonical isomorphism

$$(A \otimes B)^{\mathrm{co}} \cong B^{\mathrm{co}} \otimes A^{\mathrm{co}}$$

which can be seen for cubes and extended by taking colimits.

A category with homs enriched in **2-Cat**, with Gray's tensor product, we call a *Gray-category*: roughly speaking, this is a sesquicategory X with each homcategory $X(x, y)$ equipped with a 2-category structure, whose 2-cells are called *3-cells* of X, such that the squares (HC) have 3-cells in them, subject to appropriate axioms. Gray-categories are more general than 3-categories. In unpublished work of A. Joyal and M. Tierney, suitable algebraic models for homotopy 3-types are found to be Gray-categories in which all 1-cells, 2-cells and 3-cells are invertible.

For more details on the Gray tensor product, the interested reader should consult [Gy1, Gy2]; and, for "strong" Gray-categories, see [GPS].

7. Higher dimensions and parity complexes

Returning to cubes, we consider the case where the 3-faces do not commute. We consider the derivation scheme $\mathbf{I}[n, 3]$ given by

$$\begin{aligned}\mathbf{s}_1, \mathbf{t}_1\colon \{&\text{words } \theta \text{ of length } n \text{ in the symbols } -, 0, + \\ &\text{with precisely three 0's}\} \to \langle\langle \mathbf{FI}[n, 2] \rangle\rangle\end{aligned}$$

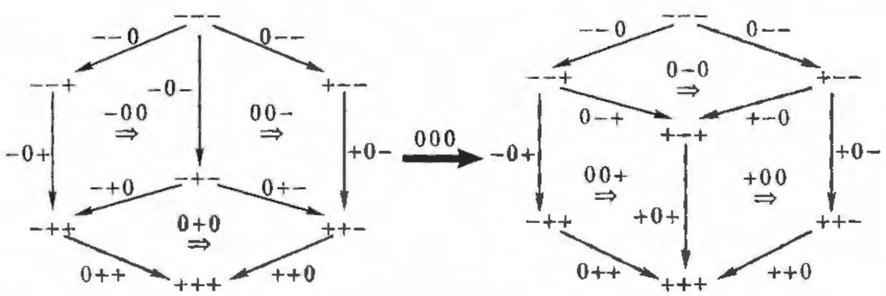
where $\mathbf{s}_1(\theta)$ = left hand side of the commuting 3-face condition, and $\mathbf{t}_1(\theta)$ = right hand side of commuting 3-face condition, for θ with 0's in the positions u, v, w. From the rewrite viewpoint, the words θ give a directed distinction between confluence checks beginning with the word obtained from θ by replacing the three 0's by $-$'s. Recall that the category $\langle\langle \mathbf{Fl}[n,2] \rangle\rangle$ has the arrows of $\mathbf{Fl}[n,2]$ as objects and has the 2-cells as arrows. While $\mathbf{Fl}[n,2]$ is a free 2-category and $\langle \mathbf{Fl}[n,2] \rangle$ is a free category, the category $\langle\langle \mathbf{Fl}[n,2] \rangle\rangle$ is not free. So $\mathbf{I}[n,3]$ is *not* a computad. It is really a "3-computad".

A *3-computad* E (where we rename graphs as "1-computads", and computads as "2-computads") is a computad $E^{\#}$ together with a derivation scheme

$$\mathbf{s}_2, \mathbf{t}_2\colon\ E_3 \to \langle\langle \mathbf{F}E^{\#} \rangle\rangle;$$

elements of E_3 are called *3-cells of E*. A *3-computad morphism* $E \to E'$ is a computad morphism $E^{\#} \to E'^{\#}$ together with a derivation scheme morphism for which the functor $\langle\langle \mathbf{F}E^{\#} \rangle\rangle \to \langle\langle \mathbf{F}E'^{\#} \rangle\rangle$ is induced by $E^{\#} \to E'^{\#}$. Each 3-computad E determines a *free 3-category* $\mathbf{F}E$. A *presentation of a 3-category* is a 3-computad together with a set of relations between 3-cells in $\mathbf{F}E$.

Here is an example of a 3-computad with one 3-cell called 0 0 0.



Each 3-cell $\theta \in E_3$ of a 3-computad E determines two 2-cells $\mathbf{s}_2(\theta)$, $\mathbf{t}_2(\theta)$ in the free 2-category $\mathbf{F}E^{\#}$. These 2-cells can be represented by string diagrams in the computad $E^{\#}$. Write θ^- for the *set* of 2-cells of $E^{\#}$ which label the nodes of a string diagram for $\mathbf{s}_2(\theta)$, and write θ^+ for the *set* of 2-cells of $E^{\#}$ which label the nodes of a string diagram for $\mathbf{t}_2(\theta)$. (These sets are independent of the choices of string diagrams in the deformation classes.) So we have two functions

$$(-)^-, (-)^+\colon\ E_3 \to \mathbf{P}(E_2)$$

where $\mathbf{P}(E_2)$ is the power set of the set of 2-cells of $E^{\#}$. In considering only the labels on nodes of a string diagram, we are, in general, disregarding quite a lot of information about the string diagram. Hence, it is a perhaps surprising consequence of the work of [S5, A1, J, S6, ASn, Pw1, Pw2] that we have:

PROPOSITION 7.1. *For 3-computads E arising from many convex polytopes such as $\mathbf{I}[n,3]$ arising from cubes, the functions $\mathbf{s}_2, \mathbf{t}_2$: $E_3 \to \langle\langle \mathbf{F}E^{\#} \rangle\rangle$ are uniquely determined by the functions $(-)^-, (-)^+$: $E_3 \to \mathbf{P}E_2$.*

At the lower dimension, the corresponding result is easily understood. For, suppose C is a (2-)computad. Then, for each 2-cell $u \in C_2$, we have paths $\mathbf{s}_1(u)$, $\mathbf{t}_1(u)$, and we can write u^-, u^+ for the *sets* of 1-cells of the graph $C^{\#}$ which occur in the respective paths. Provided the graph $C^{\#}$ has no circuits, the only other information we need to reconstruct the paths from the set is the order. However, the order is forced by knowledge of the functions $\mathbf{s}_0, \mathbf{t}_0$: $C_1 \to C_0$. So the 2-dimensional version of Proposition 7.1 is true. To be consistent at even the lowest dimension, we can define $\alpha^- = \{a\}$, $\alpha^+ = \{b\}$ for each 1-cell α: $a \to b$ of C.

In this way, each n-computad E leads to a graded set E_k, $0 \leqslant k \leqslant n$, together with functions $(-)^-, (-)^+$: $E_k \to \mathbf{P}(E_{k-1})$, $0 < k \leqslant n$. This is the basic structure involved in the higher-dimensional combinatorial notion of circuit-free graph called *parity complex* [S6, S8]; however, a parity complex is to satisfy some axioms which are not true of all such structures underlying n-computads. The axiom which somewhat reflects the source-target equations in a computad is, for all cells x of dimension $\geqslant 2$, the equality of sets

$$x^{--} \cup x^{++} = x^{-+} \cup x^{+-},$$

where the unions are disjoint, and, for example, S^- is the union of the sets x^-, $x \in S$, for any $S \subset E_k$. The main result of [S6] is the construction of the free n-category on an n-computad which is uniquely determined by the parity complex.

Following Aitchison's ideas [A2] for cubes and simplexes, we note that it is possible to use string-like diagrams to keep track of facial relations in consecutive dimensions of a parity complex. Specifically, suppose we have disjoint finite sets M, X and functions $(-)^-, (-)^+$: $M \to \mathbf{P}(X)$ such that, for all $m \neq n$ in M,

$$(m^- \cap n^-) \cup (m^+ \cap n^+) = \varnothing.$$

Put

$$\partial M = \{(-,x)\colon\ x \in X,\ x \notin M^+\} \cup \{(+,x)\colon\ x \in X,\ x \notin M^-\}.$$

Then there is a graph s, t: $X \to M \cup \partial M$ given by

$$x \in s(x)^- \cap t(x)^+ \quad \text{for } x \in M^- \cap M^+, \quad s(x) = (-,x) \quad \text{for } x \notin M^+,$$

and

$$t(x) = (+,x) \quad \text{for } x \notin M^-.$$

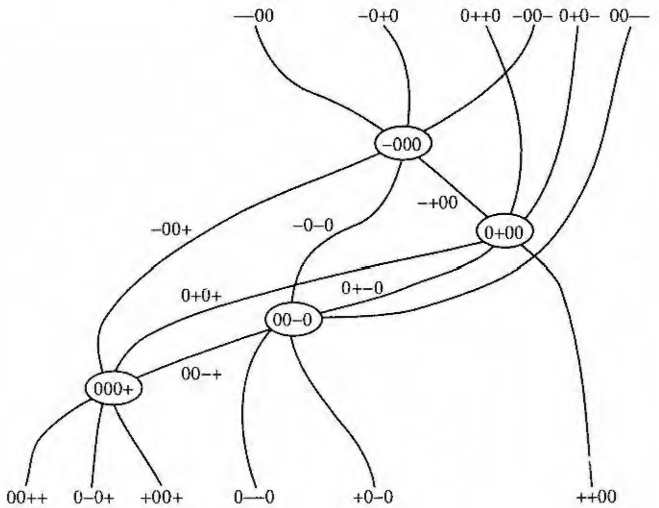
There is no reason why such a graph should be planar; however, we do draw it in the plane, with edges directed down, sometimes crossing at non-nodes, with each *inner node*

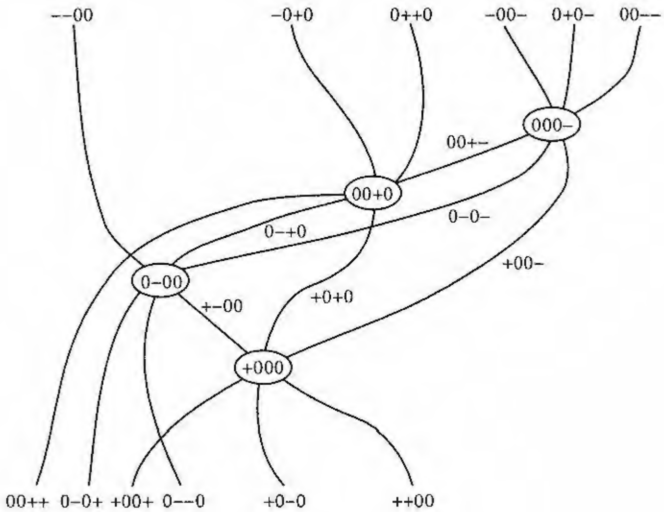
$m \in M$ labeled by m, with each *outer node* in ∂M left undistinguished, and with each edge $x \in X$ labeled by x.

Returning to cubes, we look at the 3-computad $\mathbf{I}[4,3]$. The set $\mathbf{I}[4,3]_k$ of k-cells contains the words of length 4 in the symbols $-, 0, +$ where the symbol 0 occurs precisely k times. In particular,

$$\mathbf{I}[4,3]_3 = \{-000, 0-00, 00-0, 000-, +000, 0+00, 00+0, 000+\}$$

and the parity complex structure is recorded by the string-like diagrams of [A2] as shown below.





By Proposition 7.1, each of these string-like diagrams represents a 2-cell in the 2-category **Fl**$[4,2]$. The *commuting 4-face relation* is the equality between these two 2-cells. The 3-computad **I**$[4,3]$ together with the commuting 4-face relation provides a presentation of a 3-category **Cub**$[4,3]$.

There is an explicit description of the free m-category on an m-dimensional parity complex in [S6]. In particular, there is a combinatorial model for **Fl**$[n,m]$. Except in the case $m=2$, as described above, I do not know of a combinatorial model for the n-cube **Cub**$[n,m]$ with commuting $(m+1)$-faces. Of course, we do have a presentation of the m-category **Cub**$[n,m]$ (as the m-computad **I**$[n,m]$ and the commuting m-face relations), and this suffices for many purposes.

8. The Yang–Baxter and Zamolodchikov equations

In this section we study the connection between categories and the so-called "Bazhanov–Stroganov d-simplex equations" which have arisen in statistical and quantum mechanics. We discuss here only the algebraic generic forms of these equations as found in [MN1], [MN2] where other references are provided and some of the physical significance is explained (also see [Dr1, T, JS3, JS4, Dr2, Z]):

$d=1$ *Matrix commutativity*

$$A_i^k B_k^j = B_i^k A_k^i,$$

$d=2$ *Yang–Baxter equation*

$$A_{i_1 i_2}^{k_1 k_2} B_{k_1 i_3}^{j_1 k_3} C_{k_2 k_3}^{j_2 j_3} = C_{i_2 i_3}^{k_2 k_3} B_{i_1 k_3}^{k_1 j_3} A_{k_1 k_2}^{j_1 j_2},$$

$d=3$ *Zamolodchikov equation*

$$A_{i_1 i_2 i_3}^{k_1 k_2 k_3} B_{k_1 i_4 i_5}^{j_1 k_4 k_5} C_{k_2 k_4 i_6}^{j_2 j_4 k_6} D_{k_3 k_5 k_6}^{j_3 j_5 j_6} = D_{i_3 i_5 i_6}^{k_3 k_5 k_6} C_{i_2 i_4 k_6}^{k_2 k_4 j_6} B_{i_1 k_4 k_5}^{k_1 j_4 j_5} A_{k_1 k_2 k_3}^{j_1 j_2 j_3}.$$

In these equations, observe that the subscript on a given subscript is the same as the subscript on the superscript directly above it. Also, superscripts are all j's and k's while subscripts are all k's and i's; in each case, there is a string of one letter followed by a string of the other. So the information in the equations can be recorded schematically as follows:

$$d=1: \qquad (*1)(1*) = (1*)(*1),$$

$$d=2: \qquad (*12)(1*3)(23*) = (23*)(1*3)(*12),$$

$$d=3: \qquad (*123)(1*45)(24*6)(356*) = (356*)(24*6)(1*45)(*123).$$

The symbol $*$ indicates where the letter change-over occurs. The pattern here is made clear by recording the bracketed terms on each side as rows of a matrix; this gives the formal matrix identities:

$$\begin{bmatrix} * & 1 \\ 1 & * \end{bmatrix} = \begin{bmatrix} 1 & * \\ * & 1 \end{bmatrix}$$

$$\begin{bmatrix} * & 1 & 2 \\ 1 & * & 3 \\ 2 & 3 & * \end{bmatrix} = \begin{bmatrix} 2 & 3 & * \\ 1 & * & 3 \\ * & 1 & 2 \end{bmatrix}$$

$$\begin{bmatrix} * & 1 & 2 & 3 \\ 1 & * & 4 & 5 \\ 2 & 3 & * & 6 \\ 4 & 5 & 6 & * \end{bmatrix} = \begin{bmatrix} 4 & 5 & 6 & * \\ 2 & 3 & * & 6 \\ 1 & * & 4 & 5 \\ * & 1 & 2 & 3 \end{bmatrix}$$

So the Bazhanov–Stroganov 4-simplex equation can be reconstructed from the formal matrix identity:

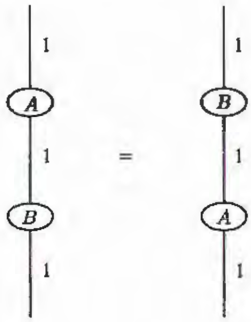
$$\begin{bmatrix} * & 1 & 2 & 3 & 4 \\ 1 & * & 5 & 6 & 7 \\ 2 & 5 & * & 8 & 9 \\ 3 & 6 & 8 & * & 10 \\ 4 & 7 & 9 & 10 & * \end{bmatrix} = \begin{bmatrix} 4 & 7 & 9 & 10 & * \\ 3 & 6 & 8 & * & 10 \\ 2 & 5 & * & 8 & 9 \\ 1 & * & 5 & 6 & 7 \\ * & 1 & 2 & 3 & 4 \end{bmatrix}$$

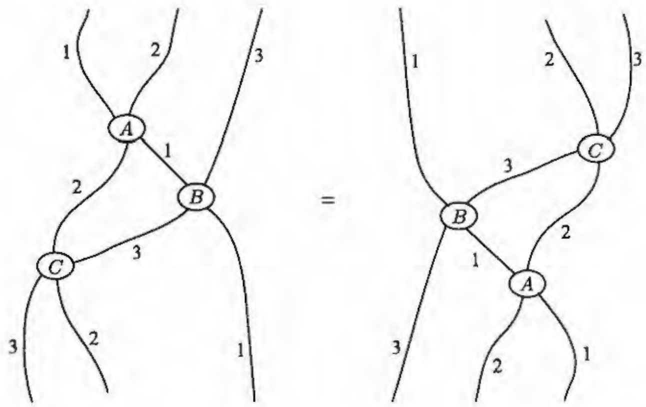
In fact, for building up these equations dimension by dimension and for dealing with the nonsquare matrices corresponding to the other entries in Aitchison's Pascal triangle of string-like diagrams, it is more convenient to renumber the strings so that the 4-simplex equation, in matrix form, becomes:

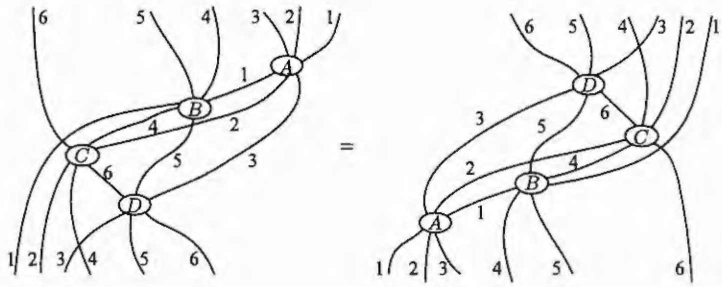
$$\begin{bmatrix} * & 1 & 2 & 4 & 7 \\ 1 & * & 3 & 5 & 8 \\ 2 & 3 & * & 6 & 9 \\ 4 & 5 & 6 & * & 10 \\ 7 & 8 & 9 & 10 & * \end{bmatrix} = \begin{bmatrix} 7 & 8 & 9 & 10 & * \\ 4 & 5 & 6 & * & 10 \\ 2 & 3 & * & 6 & 9 \\ 1 & * & 3 & 5 & 8 \\ * & 1 & 2 & 4 & 7 \end{bmatrix}$$

These formal matrices are also related to the numerical matrices for which the vanishing determinant condition [MN2] gives the dependence of the $d(d+1)/2$ parameters in the parameterized version of the d-simplex equation.

Referring to the $A, B, C, \ldots$ form of the equations, Ian Aitchisom observed (1990) that the Penrose diagrams (in the sense of [PR]) for these tensor equations occurred in his "Pascal's triangle" of string-like diagrams [A2] associated with the oriented d-cubes (*not* the d-simplexes). For $d = 2$, this reflects the well-known connections between the Yang–Baxter equation, the Coxeter relations for the symmetric groups, and paths around the edges of a cube. It should be recalled here that the ordering of the strings into, and out of, nodes is ignored (as usual with parity complexes and with Penrose notation).







Comparison with the string-like diagrams of Section 7 shows that the d-simplex equation is allied with the commuting $(d+1)$-cube.

It is possible to interpret the d-simplex equation as a morphism from a categorical structure constructed from geometry to a categorical structure of the same kind constructed from algebra. In particular, consider the Yang–Baxter equation ($d = 2$).

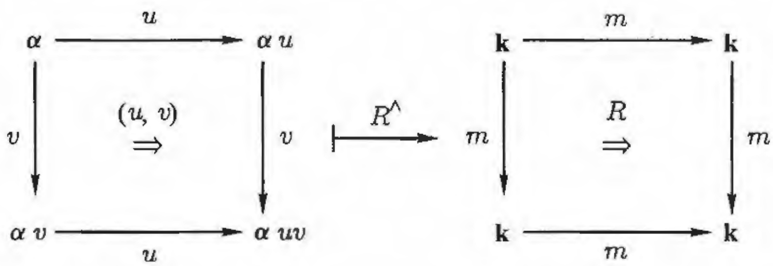
On the geometric side, recall that we have the derivation scheme $\mathbf{I}[n, 2]$ which involves the 2-dimensional faces of the n-cube; this gives the free 2-category $\mathbf{Fl}[n, 2]$.

On the algebraic side, we would like to consider a 2-category $\Sigma\mathbf{Vect}_\mathbf{k}$ whose only object is a field $\mathbf{k}$ whose arrows $V\colon\ \mathbf{k} \to \mathbf{k}$ are finite-dimensional vector spaces over $\mathbf{k}$ and whose 2-cells $t\colon\ V \Rightarrow W\colon\ \mathbf{k} \to \mathbf{k}$ are linear functions $t\colon\ V \to W$. However, we want the composition of arrows to be tensor product of vector spaces which is not strictly associative. This really provides an example of a "bicategory" which is the subject of the next section. For our present purposes, this problem can be avoided by using matrices instead of linear functions. More precisely, let $\Sigma\mathbf{Mat}_\mathbf{k}$ denote the 2-category with one object $\mathbf{k}$ whose arrows are natural numbers and whose 2-cells $A\colon\ m \Rightarrow n\colon\ \mathbf{k} \to \mathbf{k}$ are $m \times n$ matrices $A = (a_{ij})$; the vertical composition is usual multiplication of matrices, while the horizontal composite of $A\colon\ m \Rightarrow n$, $B\colon\ r \Rightarrow s$ is their Kronecker product $A \otimes B = (a_{ij} b_{pq})\colon\ mr \Rightarrow ns$.

Now suppose $R\colon\ mm \Rightarrow mm$ is a 2-cell in $\Sigma\mathbf{Mat}_\mathbf{k}$. We can extend this to a 2-functor

$$R^\wedge\colon\ \mathbf{Fl}[n, 2] \to \Sigma\mathbf{Mat}_\mathbf{k}$$

determined by the following assignment.



The matrix R is called a *Yang–Baxter matrix* when it is invertible and the 2-functor $R^\wedge$ identifies the commuting 3-face relations for some (and hence all) $n \geqslant 3$. It should be clear now how such a matrix R provides a solution to the Yang–Baxter equation. There is an induced 2-functor $R^\wedge\colon\ \mathbf{Cub}[n, 2] \to \Sigma\mathbf{Mat}_\mathbf{k}$.

Now consider applying the same ideas to the Zamolodchikov equation. On the geometric side there is no problem since we have the free 3-category $\mathbf{Fl}[n, 3]$. A small difficulty arises on the algebraic side when we try to push the category of vector spaces up another dimension. This time we would like to consider a 3-category $\Sigma^2\mathbf{Vect}_\mathbf{k}$ wbose ouly object is a field $\mathbf{k}$ whose only arrow is the identity of $\mathbf{k}$ whose 2-cells V are finite-dimensional vector spaces over $\mathbf{k}$ and whose 3-cells $t\colon\ V \to M$ are linear functions. This time two of the compositions are to be tensor product with the third taken to be composition of linear functions, as before. The problem of nonstrictness of associativity of tensor product can be avoided as before by using matrices, however, now we also require the middle-four-interchange law:

$$(U \otimes V) \otimes (W \otimes X) = (U \otimes W) \otimes (V \otimes X)$$

which of course does not strictly hold; there is only a canonical isomorphism in place of the equality. This problem cannot be avoided. In fact, $\Sigma^2\mathbf{Vect_k}$ is an example of a *tricategory* in the sense of [GPS]. Using matrices, we obtain a Gray-category $\Sigma^2\mathbf{Mat_k}$. (It is shown in [GPS] that, more generally, every tricategory is "triequivalent" to a Gray-category.) As we mentioned at the end of Section 6, every 3-category is a Gray-category. It is therefore meaningful to consider Gray-functors from a 3-category to a Gray-category. In particular, each $m^3 \times m^3$ matrix R induces such a Gray-functor

$$R^\wedge\colon\ \mathbf{Fl}[n,3] \to \Sigma^2\mathbf{Mat_k}$$

we call R a *Zamolodchikov matrix* when it is invertible and $R^\wedge$ identifies the commuting 4-face relations for some (and hence all) $n \geqslant 4$. Such a matrix R provides a solution to the Zamolodchikov equation.

Higher dimensions offer no new problems. For the d-simplex equation, there is an appropriate structure $\Sigma^{d-1}\mathbf{Mat_k}$ with precisely one i-cell for each $i \leqslant d-2$, whose $(d-1)$-cells are natural numbers, whose d-cells are matrices, whose first $d-1$ compositions are Kronecker product (among which the middle-four-interchange law holds only up to a coherent invertible d-cell), and whose remaining composition is usual matrix product (which strictly satisfies the middle-four-interchange law with each earlier composition). A *d-simplex matrix* is an invertible $m^d \times m^d$ matrix R which induces a structure-preserving morphism

$$R^\wedge\colon\ \mathbf{Cub}[d+1,d] \to \Sigma^{d-1}\mathbf{Mat_k}.$$

9. Bicategories

Bicategories (and the appropriate 3-graph with bicategories as 0-cells) were first defined by Bénabou [Bn1, Bn2].

A *bicategory* $\mathbf{B}$ is a 2-graph equipped with the following extra structure:

(Ba) for each pair of objects a, b, a category structure on the graph $\mathbf{B}(a,b)$ with composition called *vertical* and denoted by $\bullet$ ("invertibility" for 2-cells will mean with respect to this composition);

(Bb) for each triple of objects a, b, c a functor

$$\circ\colon\ \mathbf{B}(a,b) \times \mathbf{B}(b,c) \to \mathbf{B}(a,c),$$

called the *horizontal composition* and written between the arguments;

(Bc) for each object a, an arrow $1_a\colon\ a \to a$ called the *identity* for a;

(Bd) invertible 2-cells

$$a_{\alpha,\beta,\gamma}\colon\ \alpha\circ(\beta\circ\gamma)\Rightarrow(\alpha\circ\beta)\circ\gamma\colon\ a\to d,$$

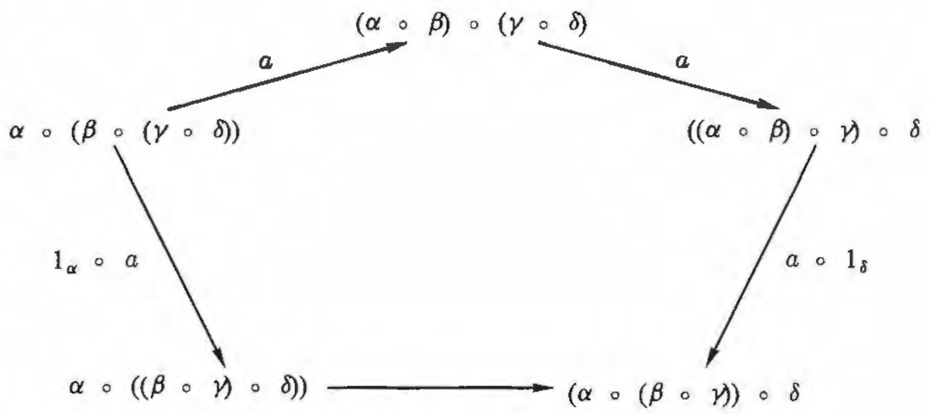
called *associativity constraints*, which are natural in $\alpha,\beta,\gamma\in\mathbf{B}(a,b)\times\mathbf{B}(b,c)\times\mathbf{B}(c,d)$;

(Be) invertible 2-cells

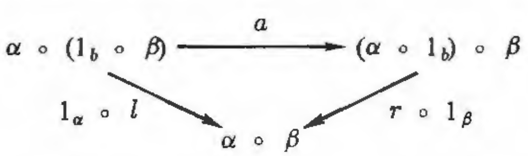
$$l_\alpha\colon\ 1_a\circ\alpha\Rightarrow\alpha\colon\ a\to b,\quad r_\alpha\colon\ \alpha\circ 1_b\Rightarrow\alpha\colon\ a\to b,$$

called *identity constraints*, which are natural in $\alpha\in\mathbf{B}(a,b)$;
subject to the following commutativity conditions:

(B1) *pentagon for associativity constraints*



(B2) *triangle for identity constraints*



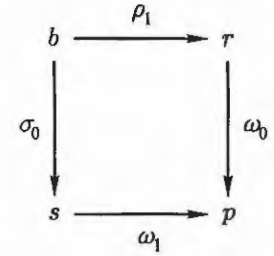
EXAMPLE 1. Let A be a category equipped with a choice of pushouts for each pair of arrows with common source. There is a bicategory Cospn A defined as follows. The objects are the objects of A. For $a,b\in A$, the category $(\mathbf{Cospn}\,A)(a,b)$ has as objects triples (ρ_0,t,ρ_1), called *cospans from a to b*, consisting of an object r and arrows $\rho_0\colon\ a\to r$, $\rho_1\colon\ b\to r$ of A; an arrow $\phi\colon\ (\rho_0,r,\rho_1)\to(\sigma_0,s,\sigma_1)$ of cospans is an arrow $\phi\colon\ r\to s$ in A such that

$$\rho_0\circ\phi=\sigma_0,\quad \rho_1\circ\phi=\sigma_1.$$

Given $(\rho_0, r, \rho_1) \in (\mathrm{Cospn}\, A)(a, b)$ and $(\sigma_0, s, \sigma_1) \in (\mathrm{Cospn}\, A)(b, c)$, define

$$(\rho_0, r, \rho_1) \circ (\sigma_0, s, \sigma_1) = (\rho_0 \circ \omega_0, p, \sigma_1 \circ \omega_1)$$

where the square



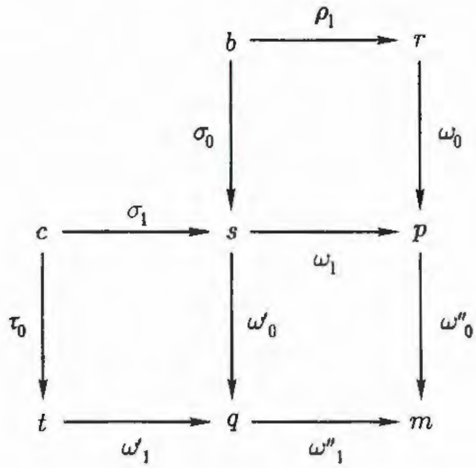
is the selected pushout of ρ_0, σ_1. Using the universal property of pushout, we can extend this functorially to arrows of spans as required for (Bb). The identity cospan for a is $1_a = (1_a, a, 1_a)$. Given cospans

$$(\rho_0, r, \rho_1) \in (\mathrm{Cospn}\, A)(a, b), (\sigma_0, s, \sigma_1) \in (\mathrm{Cospn}\, A)(b, c)$$

and

$$(\tau_0, t, \tau_1) \in (\mathrm{Cospn}\, A)(c, d),$$

we can form the diagram



of selected pushouts. Each of the cospans

$$(\rho_0, r, \rho_1) \circ ((\sigma_0, s, \sigma_1) \circ (\tau_0, t, \tau_1)), \quad ((\rho_0, r, \rho_1) \circ (\sigma_0, s, \sigma_1)) \circ (\tau_0, t, \tau_1)$$

is canonically isomorphic to $(\rho_0 \circ \omega_0 \circ \omega_0'', m, \rho_1 \circ \omega_1 \circ \omega_1'')$, and so, to each other, yielding the associativity constraints. There are also canonical isomorphisms

$$1_a \circ (\rho_0, r, \rho_1) \cong (\rho_0, r, \rho_1) \cong (\rho_0, r, \rho_1) \circ 1_b$$

yielding the identity constraints. To check commutativity of (B1), (B2), it suffices to check after composition with the coprojections ω_0, ω_1 into the appropriate pushouts, and we recommended this as an exercise.

EXAMPLE 2. A monoidal (= "tensor") category **V** (in the sense of [EK]) can be defined to be a bicategory **B** with one object. More precisely, if **V** is a monoidal category then a bicategory with only one object a is defined by $\mathbf{B}(a, a) = \mathbf{V}$; the tensor product of **V** is the horizontal composition $\circ$ of **B**. While, if a is any object of a bicategory **B** then $\mathbf{B}(a, a)$ becomes a monoidal category. For many purposes it is convenient to distinguish **V** from the one-object **B**; the notation $\Sigma\mathbf{V}$ for **B** is not bad.

EXAMPLE 3. A bicategory in which all the constraints are identities in a 2-category (Section 2). As each category A can be regarded as a 2-category for which each category $A(a, b)$ is discrete, we can also regard categories as special bicategories.

EXAMPLE 4. There is a bicategory **Prof** which stands in relation to the 2-category **Cat** much as the category of sets and relations stands in relation to the category **Set** of sets and functions. The objects of **Prof** are categories. An arrow $M\colon A \nrightarrow B$ is a *profunctor* (also called "distributor" [Bn2], "bimodule" [L], or just "module" [S3]); that is, a functor $M\colon A^{\mathrm{op}} \times B \to \mathbf{Set}$. The 2-cells $M \Rightarrow N$ are natural transformations between the functors. Composition of profunctors $M\colon A \nrightarrow B$, $N\colon A \nrightarrow B$ is given by the coend formula (see [ML1, ML2] for the history of "ends"):

$$(M \circ N)(a, c) = \int^b M(a, b) \times N(b, c).$$

Suppose $\mathbf{B}, \mathbf{X}$ are bicategories. A *lax functor* (also called "morphism of bicategories")

$$T\colon \mathbf{B} \to \mathbf{X}$$

is a 2-graph morphism which is functorial on vertical composition and is equipped with the following extra structure:

(LFa) for each object a of **B**, a 2-cell $i_a\colon 1_{T(a)} \Rightarrow T(1_a)$ of **X**;

(LFb) 2-cells $m_{\alpha,\beta}\colon T(\alpha) \circ T(\beta) \Rightarrow T(\alpha \circ \beta)$ which are natural in $(\alpha, \beta) \in \mathbf{B}(a, b) \times \mathbf{B}(b, c)$; subject to the following commutativity conditions:

(LF1)

$$\begin{array}{ccccc}
 & (T(\alpha)\circ T(\beta))\circ T(\gamma) & \xrightarrow{\;m\,\circ\,1\;} & T(\alpha\circ\beta)\circ T(\gamma) & \\
{\scriptstyle a}\nearrow & & & & \searrow{\scriptstyle m} \\
(T(\alpha)\circ(T(\beta)\circ T(\gamma)) & & & & T((\alpha\circ\beta)\circ\gamma) \\
{\scriptstyle 1\circ m}\searrow & & & & \nearrow{\scriptstyle T_a} \\
 & T(\alpha)\circ T(\beta\circ\gamma) & \xrightarrow[\;m\;]{} & T(\alpha\circ(\beta\circ\gamma)) &
\end{array}$$

(LF2)

$$\begin{array}{ccc}
1_{T(a)}\circ T(\alpha) & \xrightarrow{\;l\;} & T(\alpha) \\
{\scriptstyle i\circ 1}\big\downarrow & & \big\uparrow{\scriptstyle T(l)} \\
T(1_a)\circ T(\alpha) & \xrightarrow[\;m\;]{} & T(1_a\circ\alpha)
\end{array}
\qquad
\begin{array}{ccc}
T(\alpha)\circ 1_{T(b)} & \xrightarrow{\;r\;} & T(\alpha) \\
{\scriptstyle 1\circ i}\big\downarrow & & \big\uparrow{\scriptstyle T(r)} \\
T(\alpha)\circ T(1_b) & \xrightarrow[\;m\;]{} & T(\alpha\circ 1_b)
\end{array}$$

EXAMPLE 5. Suppose $F\colon A\to X$ is a functor between categories with selected pushouts. Then there is a lax functor $T=\mathrm{Cospn}(F)\colon \mathrm{Cospn}\,A\to\mathrm{Cospn}\,X$ described as follows. Let T take a general 2-cell $\phi\colon (\rho_0,r,\rho_1)\Rightarrow(\sigma_0,s,\sigma_1)\colon a\to b$ in $\mathrm{Cospn}\,A$ to the 2-cell

$$F(\phi)\colon \big(F(\rho_0),F(r),F(\rho_1)\big)\Rightarrow\big(F(\sigma_0),F(s),F(\sigma_1)\big)\colon F(a)\to F(b)$$

in $\mathrm{Cospn}\,X$. The 2-cells of (LFa) are identities. The universal property of pushouts in X yields a canonical comparison arrow from the pushout of $F(\rho_0)$, $F(\sigma_1)$ to $F(p)$ (in the notation of Example 1). This gives the data for (LFb). The axioms (LF1), (LF2) are easily verified.

EXAMPLE 6. A monoidal functor $F\colon \mathbf{V}\to\mathbf{W}$ (in the sense of [EK]) amounts precisely to a lax functor $T\colon \Sigma\mathbf{V}\to\Sigma\mathbf{W}$ (see Example 2).

EXAMPLE 7. For bicategories $\mathbf{B},\mathbf{X}$, a lax functor $T\colon \mathbf{B}\to\mathbf{X}$ is called a *pseudo functor* (also called "homomorphism" in [Bn1, Bn2]) when all the 2-cells

$$m_{\alpha,\beta}\colon T(\alpha)\circ T(\beta)\Rightarrow T(\alpha\circ\beta)$$

and i_a: $1_{T(a)} \Rightarrow T(1_a)$ are invertible. When these 2-cells are all identities, T is called a *2-functor*; when **B**, **X** are both 2-categories (see Example 3) this agrees with the terminology in Section 2. It is perhaps of interest that, for any category C, pseudo functors T: $C^{\text{op}} \to \mathbf{Cat}$ are equivalent, via the "Grothendieck construction", to functors P: $E \to C$ which are *fibrations*; in particular, when C is a group (regarded as a category with one object and all arrows invertible), such a T is a Schreier *factor system* as occur in group cohomology (for example, see [Gd]). A lax functor T: $\mathbf{B} \to \mathbf{X}$ is called *normalized* when all the 2-cells i_a: $1_{T(a)} \Rightarrow T(a_1)$ are identities. Jean Bénabou [Bn2] has shown how to construct, from *every* functor (not just fibrations!) P: $E \to C$, a normalized lax functor T: $C^{\text{op}} \to \mathbf{Prof}$ (see Example 4); the Grothendieck construction generalizes to reverse this construction.

EXAMPLE 8. Let **1** denote the one-object discrete category. A lax functor T: $\mathbf{1} \to \mathbf{B}$ amounts to a monad in **B** (also see Section 5).

EXAMPLE 9. Lax functors can be composed in a fairly obvious way (which we leave to the reader) yielding a category **Bicat** whose objects are bicategories and whose arrows are lax functors.

EXAMPLE 10. Each object k of a bicategory **B** determines a pseudo functor

$$H_k = \mathbf{B}(k, -)\colon\ \mathbf{B} \to \mathbf{Cat}$$

called the *pseudo functor represented by k*. The category $H_k(a)$ is $\mathbf{B}(k, a)$. The functor $H_k(\alpha)$: $\mathbf{B}(k,a) \to \mathbf{B}(k,b)$ is given by composing on the right with α: $a \to b$. For each 2-cell u: $\alpha \Rightarrow \beta$, the natural transformation $H_k(u)$: $H_k(\alpha) \Rightarrow H_k(\beta)$ has component $H_k(u)_\varepsilon = \varepsilon \circ u$: $\varepsilon \circ \alpha \Rightarrow \varepsilon \circ \beta$ at $\varepsilon \in \mathbf{B}(k, a)$. The natural transformation

$$i_a\colon\ 1_{\mathbf{B}(k,a)} \Rightarrow - \circ 1_a$$

is provided by the inverse of the identity constraint r. The natural transformation

$$m_{\alpha,\beta}\colon\ H_k(\alpha) \circ H_k(\beta) \Rightarrow H_k(\alpha \circ \beta)$$

has component $(\varepsilon \circ \alpha) \circ \beta \to \varepsilon \circ (\alpha \circ \beta)$ at $\varepsilon \in \mathbf{B}(k, a)$ given by the inverse of the associativity constraint a. Axiom (LF1) is a pentagon since **Cat** is a 2-category and it amounts to axiom (B1) for **B**. We leave (LF2) as an exercise.

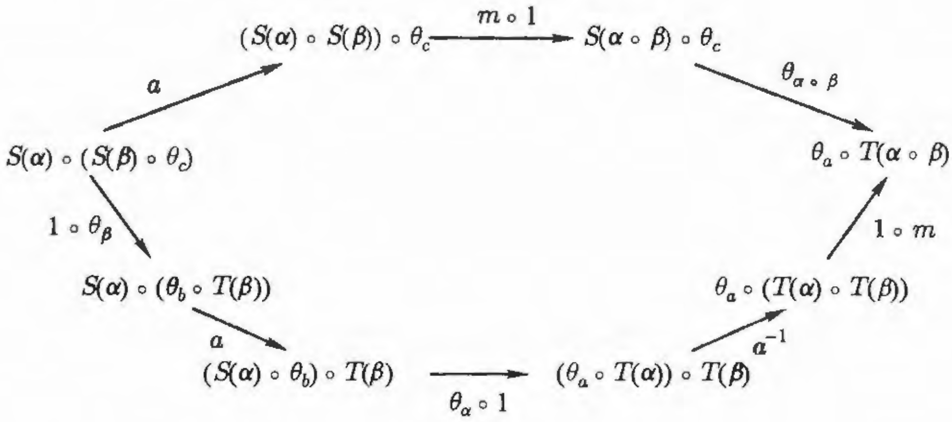
Suppose S, T: $\mathbf{B} \to \mathbf{X}$ are lax functors. A *transformation* θ: $S \Rightarrow T$ consists of the following data:

(Ta) for each object a of **B**, an arrow θ_a: $S(a) \to T(a)$ of **X**;

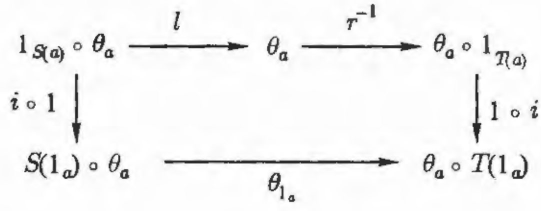
(Tb) 2-cells θ_α: $S(\alpha) \circ \theta_b \Rightarrow \theta_a \circ T(\alpha)$ which are natural in $\alpha \in \mathbf{B}(a, b)$;

such that the following commutativity conditions hold:

(T1)



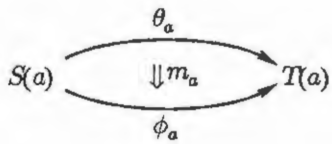
(T2)



A transformation θ: $S \Rightarrow T$ is called *strong* when each of the 2-cells θ_α: $S(\alpha) \circ \theta_b \Rightarrow \theta_a \circ T(\alpha)$ is invertible.

EXAMPLE 11. Suppose κ: $h \to k$ is an arrow of a bicategory $\mathbf{B}$. There is a strong transformation $\theta = H_\kappa$: $H_k \Rightarrow H_h$ whose component θ_a: $\mathbf{B}(k, a) \to \mathbf{B}(h, a)$ is the functor given by composition on the left with κ, and whose natural isomorphism θ_α: $H_k(\alpha) \circ \theta_b \Rightarrow \theta_a \circ H_h(\alpha)$ has component at ξ: $k \to a$ given by the associativity constraint a: $\kappa \circ (\xi \circ \alpha) \to (\kappa \circ \xi) \circ \alpha$.

Suppose θ, ϕ: $S \Rightarrow T$: $\mathbf{B} \to \mathbf{X}$ are transformations. A *modification* m: $\theta \to \phi$ is a family of 2-cells



subject to the following commutativity condition:

(M)

$$\begin{array}{ccc} S(\alpha) \circ \theta_b & \xrightarrow{\theta_\alpha} & \theta_a \circ T(\alpha) \\ {\scriptstyle 1 \circ m_b} \downarrow & & \downarrow {\scriptstyle m_a \circ 1} \\ S(\alpha) \circ \phi_b & \xrightarrow[\phi_\alpha]{} & \phi_a \circ T(\alpha) \end{array}$$

EXAMPLE 12. Each 2-cell $w\colon \kappa \Rightarrow \lambda\colon k \to h$ in a bicategory $\mathbf{B}$ yields a modification

$$H_w\colon H_\kappa \to H_\lambda\colon H_k \Rightarrow H_h\colon \mathbf{B} \to \mathbf{Cat}$$

whose component at $a \in \mathbf{B}$ is the natural transformation given by horizontal composition on the left with the 2-cell w.

Modifications $m\colon \theta \to \phi$, $n\colon \phi \to \psi$ can be composed to yield a modification $m \bullet n\colon \theta \to \psi$ using pointwise vertical composition in $\mathbf{X}$. Transformations $\theta\colon S \Rightarrow T$, $\theta'\colon T \Rightarrow U$ can be composed to yield a transformation $\theta \circ \theta'\colon S \Rightarrow U$ by putting

$$(\theta \circ \theta')_a = \theta_a \circ \theta'_a$$

and

$$(\theta \circ \theta')_\alpha = \left(\begin{array}{c} S(\alpha) \circ (\theta_b \circ \theta'_b) \xrightarrow{a} \big(S(\alpha) \circ \theta_b\big) \circ \theta'_b \xrightarrow{\theta_\alpha \circ 1} \big(\theta_a \circ T(\alpha)\big) \circ \theta'_b \\ \xrightarrow{a^{-1}} \theta_a \circ \big(T(\alpha) \circ \theta'_b\big) \xrightarrow{1 \circ \theta'_\alpha} \theta_a \circ \big(\theta'_a \circ U(\alpha)\big) \xrightarrow{a} (\theta_a \circ \theta'_a) \circ U(\alpha) \end{array} \right);$$

this composition is not strictly associative, but the associativity and identity constraints of $\mathbf{X}$ yield associativity and identity constraints here. This describes a bicategory $\mathbf{Lax}(\mathbf{B}, \mathbf{X})$ whose objects are lax functors, whose arrows are transformations, and whose 2-cells are modifications. Write $\mathbf{Psd}(\mathbf{B}, \mathbf{X})$ for the subbicategory of $\mathbf{Lax}(\mathbf{B}, \mathbf{X})$ consisting of the pseudo functors $T\colon \mathbf{B} \to \mathbf{X}$, the strong transformations between these, and the modifications between these. Notice that $\mathbf{Lax}(\mathbf{B}, \mathbf{X})$ and $\mathbf{Psd}(\mathbf{B}, \mathbf{X})$ are 2-categories if $\mathbf{X}$ is a 2-category (there is no need for $\mathbf{B}$ to be).

EXERCISE. Show that a lax functor $\mathbf{1} \to \mathbf{Lax}(\mathbf{1}, \mathbf{X})$ amounts to a pair of monads on the same object of $\mathbf{X}$ together with a distributive law between the monads (see Section 5).

For each bicategory $\mathbf{B}$, there is a pseudo functor

$$\mathcal{Y}\colon \mathbf{B} \to \mathbf{Psd}(\mathbf{B}, \mathbf{Cat})^{\mathrm{op}}$$

the letter "Y" is for Yoneda since this is a generalization of the Yoneda embedding of categories. The value of $\mathcal{Y}$ at a 2-cell w: $\kappa \Rightarrow \lambda$: $k \to h$ in **B** is the displayed modification in Exercise 11. The data (LFa), (LFb) for $\mathcal{Y}$ are supplied by the identity and associativity constraints of **B**.

For any pseudo functor T: $\mathbf{B} \to \mathbf{Cat}$, we shall describe a strong transformation

$$\mathbf{e}\colon \mathbf{Psd}(\mathbf{B}, \mathbf{Cat})(\mathcal{Y}, T) \Rightarrow T\colon \mathbf{B} \to \mathbf{Cat}.$$

For each $k \in \mathbf{B}$, the functor $\mathbf{e}_k$: $\mathbf{Psd}(\mathbf{B}, \mathbf{Cat})(H_k, T) \to T(k)$ takes an arrow m: $\theta \to \phi$ in the category $\mathbf{Psd}(\mathbf{B}, \mathbf{Cat})(H_k, T)$ to the arrow $m_k(1_k)$: $\theta_k(1_k) \to \phi_k(1_k)$ in the category $T(k)$. For each κ: $k \to h$ in **B**, the natural isomorphism

$$\mathbf{e}_\kappa\colon H_\kappa \circ \mathbf{e}_h \Rightarrow \mathbf{e}_k \circ T(\kappa)\colon \mathbf{Psd}(\mathbf{B}, \mathbf{Cat})(H_k, T) \to T(h)$$

whose component at the object θ of $\mathbf{Psd}(\mathbf{B}, \mathbf{Cat})(H_k, T)$ is the isomorphism

$$\theta_\kappa\colon H_k(\kappa) \circ \theta_h \Rightarrow \theta_k \circ T(\kappa).$$

PROPOSITION 9.1 (Bicategorical Yoneda Lemma [S3]). *For each object k of the bicategory* **B** *and each pseudo functor T:* $\mathbf{B} \to \mathbf{Cat}$*, the functor*

$$\mathbf{e}_k\colon \mathbf{Psd}(\mathbf{B}, \mathbf{Cat})(H_k, T) \to T(k)$$

is an equivalence of categories.

An arrow α: $a \to b$ in a bicategory **B** is called an *equivalence* when there exist an arrow β: $b \to a$ and invertible 2-cells $\alpha \circ \beta \Rightarrow 1_a$, $1_b \Rightarrow \beta \circ \alpha$; write α: $a \xrightarrow{\sim} b$. For example, using the axiom of choice, one can see that an arrow f: $A \to B$ in **Cat** is an equivalence if and only if the functor f: $A \to B$ is full, faithful and each object b of B is isomorphic to an object of the form $f(a)$ for some $a \in A$. As another example, an arrow θ in $\mathbf{Psd}(\mathbf{B}, \mathbf{X})$ is an equivalence if and only if each arrow θ_a is an equivalence in **X**.

Hence, the bicategorical Yoneda lemma states that **e** is an equivalence in the bicategory $\mathbf{Psd}(\mathbf{B}, \mathbf{Cat})$. Notice that $\mathcal{Y}$ and hence $\mathbf{Psd}(\mathbf{B}, \mathbf{Cat})(\mathcal{Y}, T)$ are 2-functors if **B** is a 2-category, so we obtain the following result which is an example of a "coherence theorem".

COROLLARY 9.2. *If* **B** *is a* 2*-category then every pseudo functor T:* $\mathbf{B} \to \mathbf{Cat}$ *is equivalent, in the* 2*-category* $\mathbf{Psd}(\mathbf{B}, \mathbf{Cat})$*, to a* 2*-functor.*

A lax functor T: $\mathbf{B} \to \mathbf{X}$ is called a *biequivalence* when it is a pseudo functor, each of the functors T: $\mathbf{B}(a, b) \to \mathbf{X}(T(a), T(b))$ is an equivalence, and, for each object x of **X**, there exists an object a of **B** and an equivalence $T(a) \xrightarrow{\sim} x$ in **X**. Using the axiom of choice, we can see that T: $\mathbf{B} \to \mathbf{X}$ is a biequivalence if and only if there exists a lax functor S: $\mathbf{X} \to \mathbf{B}$ and equivalences $T \circ S \xrightarrow{\sim} 1_\mathbf{B}$, $1_\mathbf{X} \xrightarrow{\sim} S \circ T$ in the bicategories $\mathbf{Lax}(\mathbf{B}, \mathbf{B})$, $\mathbf{Lax}(\mathbf{X}, \mathbf{X})$, respectively.

The following proof is due to R. Gordon and A.J. Power and was made public at the 1991 Summer Category Theory Conference in Montréal.

PROPOSITION 9.3 [MP]. *For every bicategory* **B**, *there exists a 2-category* **K** *with a biequivalence* $\mathbf{B} \to \mathbf{K}$.

PROOF. It follows from the bicategorical Yoneda lemma that the functors

$$\mathcal{Y}\colon\ \mathbf{B}(a,b) \to \mathbf{Psd}(\mathbf{B},\mathbf{Cat})^{\mathrm{op}}(H_a, H_b)$$

are equivalences. So we can take **K** to be the sub-2-category of $\mathbf{Psd}(\mathbf{B},\mathbf{Cat})^{\mathrm{op}}$ obtained by restricting to those objects of the form H_a. Then $\mathcal{Y}$ gives the desired biequivalence. □

A direct proof, based on the above recall (Example 2), that every monoidal category is monoidally equivalent to a strict monoidal category, can be found in [JS5]. The result [GPS] for the next dimension is that every tricategory is "triequivalent" to a Gray-category (not in general to a 3-category). These references also explain how to extract from this result the coherence theorems in the more familiar form "all diagrams commute".

10. Nerves

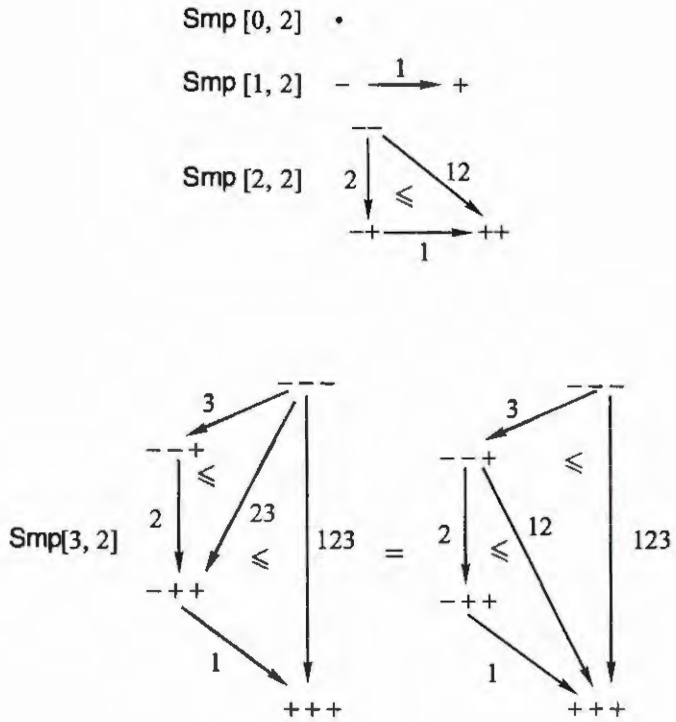
The purpose of forming the nerve of a categorical structure is to create an object which contains all tbe information of the structure and yet is in a form more able to be compared with familiar geometric structures. There is a notion of cubical nerve, but we shall deal with the more usual simplicial nerve. In preparation for this, we need to modify our discussion of cubes to extract simplexes. For each natural number r, consider the word $\alpha_{r,n}$ of length n in the symbols $-,+$ which begins with r minuses and ends with $n-r$ pluses.

$$\alpha_{r,n} = \underbrace{- - \cdots -}_{r}\underbrace{+ + \cdots +}_{n-r}$$

Let $\mathsf{Smp}[n,m]$ denote the sub-m-category of $\mathsf{Cub}[n,m]$ obtained by taking only the objects $\alpha_{r,n}$. The m-category $\mathsf{Smp}[n,m]$ is the *n-simplex with commuting* $(m+1)$-*faces*. (There is an analogue of Proposition 4.1.) In particular, $\mathsf{Smp}[n,1]$ is a linearly ordered set with $n+1$ elements; it is more usual to use the ordered set

$$[n] = \{0, 1, \ldots, n\}.$$

Also, we have the 2-categories (using "position" notation):



Recall that ⟨**Cat**⟩ denotes the category of (small) categories and functors. The category $\boldsymbol{\Delta}$ of finite nonempty ordinals and order-preserving functions is the full subcategory $\boldsymbol{\Delta}$ of ⟨**Cat**⟩ consisting of the categories $[n]$. A *simplicial set* is a functor S: $\boldsymbol{\Delta}^{\mathrm{op}} \to \mathbf{Set}$; its value at $[n]$ is denoted by S_n. The *nerve* $N(A)$ of a category A is the simplicial set obtained by restricting the representable functor

$$\langle\mathbf{Cat}\rangle(-, A)\colon\ \langle\mathbf{Cat}\rangle^{\mathrm{op}} \to \mathbf{Set} \text{ to } \boldsymbol{\Delta}^{\mathrm{op}};$$

so

$$N(A)_n = \langle\mathbf{Cat}\rangle\big([n], A\big).$$

This construction is obviously functorial in $A \in \langle\mathbf{Cat}\rangle$, so we obtain nerve as a functor

$$N\colon\ \langle\mathbf{Cat}\rangle \to [\boldsymbol{\Delta}^{\mathrm{op}}, \mathbf{Set}]$$

into the category $[\boldsymbol{\Delta}^{\mathrm{op}}, \mathbf{Set}]$ of simplicial sets. It is easily seen that this functor is full, faithful, and has a left adjoint which preserves finite products. The simplicial sets S

which are isomorphic to nerves of categories can be characterized as those functors $S\colon \boldsymbol{\Delta}^{\mathrm{op}} \to \mathbf{Set}$ which preserve pullbacks; but they can also be characterized as those S for which each *admissible horn* has a unique filler (see [S4, S5, S7] for this terminology).

There is a canonical 2-functor $\mathsf{Smp}[n,2] \to \mathsf{Smp}[n,1]$ which is the identity function on objects and identifies the 2-cells. Each functor $f\colon \mathsf{Smp}[n,1] \to \mathsf{Smp}[n',1]$ has a lifting to a 2-functor $f'\colon \mathsf{Smp}[n,2] \to \mathsf{Smp}[n',2]$ uniquely determined by the condition that each arrow $f'(r\colon \alpha_{r,n} \to \alpha_{r+1,n})$ is given by the natural ordering of $f(\alpha_{r,n})\backslash f(\alpha_{r+1,n})$. This gives a functor

$$j\colon \boldsymbol{\Delta} \to \langle \mathbf{2\text{-}Cat} \rangle, \quad [n] \mapsto \mathsf{Smp}[n,2], \ f \mapsto f'.$$

The *nerve* $N(K)$ of a 2-category K is the simplicial set obtained by composing the functor $j^{\mathrm{op}}\colon \boldsymbol{\Delta}^{\mathrm{op}} \to \langle \mathbf{2\text{-}Cat} \rangle^{\mathrm{op}}$ with the representable functor

$$\langle \mathbf{2\text{-}Cat} \rangle(-,K)\colon \langle \mathbf{2\text{-}Cat} \rangle^{\mathrm{op}} \to \mathbf{Set}.$$

So, an element of $N(K)$ of dimension n is a 2-functor $x\colon \mathsf{Smp}[n,2] \to K$; we think of this as an n-simplex in K with commuting 3-faces. We obtain a nerve functor

$$N\colon \langle \mathbf{2\text{-}Cat} \rangle \to [\boldsymbol{\Delta}^{\mathrm{op}}, \mathbf{Set}]$$

with a left adjoint; but this time the functor is not full. We need to take account of more structure on the simplicial set $N(K)$, namely, those elements of dimension 2 which are *commutative* triangles. It is possible [S4] to characterize (up to isomorphism) nerves of 2-categories as simplicial sets, with some distinguished elements (called "hollow" or "thin"), satisfying some axioms the main one of which states that each admissible horn should have a unique thin filler.

There is also a notion of nerve for a bicategory [DS] which has not received much attention. Let $\mathbf{Bicat}_{\mathbf{norm}}$ denote the category whose objects are bicategories and whose arrows are normalized lax functors. As every category is a bicategory, we can regard $\boldsymbol{\Delta}$ as a subcategory of $\mathbf{Bicat}_{\mathbf{norm}}$. For each bicategory $\mathbf{B}$, the composite of the inclusion of $\boldsymbol{\Delta}^{\mathrm{op}}$ in $\mathbf{Bicat}_{\mathbf{norm}}(-,\mathbf{B})$ with the representahle

$$\mathbf{Bicat}_{\mathbf{norm}}(-,\mathbf{B})\colon \mathbf{Bicat}_{\mathbf{norm}}^{\mathrm{op}} \to \mathbf{Set}$$

is defined to be the *nerve* $N(\mathbf{B})$ of $\mathbf{B}$; so

$$N(\mathbf{B})_n = \mathbf{Bicat}_{\mathbf{norm}}\big([n],\mathbf{B}\big).$$

EXERCISE. For a 2-category K, the nerve of K as a 2-category is isomorphic to the nerve of K as a bicategory.

EXERCISE. Biequivalent bicategories have homotopically equivalent nerves. (See [GZ] for homotopy for simplicial sets.)

The nerve of an m-category was made precise in [S5], and other approaches appear in [A1, JW, ASn]. Essentially each proceeds as above after giving a precise description of $\mathbf{Smp}[n,m]$. Verity [V] has shown that this nerve functor, defined on $\langle \mathbf{m\text{-}Cat} \rangle$ and viewed as landing in the category of simplicial sets with distinguished "hollow" (or "thin") elements, is fully faithful. A good deal of progress has been made by Michael Zaks and Dominic Verity on the characterization (up to isomorphism) of these nerves; but at the time of writing (November 1992), the conjecture of John Roberts (see [S5]) remains unproved.

Finally, we remark that categorical structures can be considered inside categories whose objects are more geometric than sets. Nerves then are simplicial geometric objects whose "geometric realizations" are "classifying spaces" [Sg].

Acknowledgement

While this paper is really neither a survey paper nor a joint paper, it contains my version of ideas from several collaborations and discussions with Iain Aitchison, Samuel Eilenberg, Michael Johnson, and Steve Schanuel.

Added in proof

This paper was completed in November 1992. The references have been updated during proofreading and [Sl, S8, S9] have been added. We point to [S9] as suitable for further reading in the area.

There have been two notable developments in the last three years. In July 1993, Dominic Verity completed the proof of the Roberts conjecture (see the end of Section 10). Also, Verity and the author have developed the use of surface diagrams for tricategories generalising the use of string diagrams for bicategories.

References

[SGA] M. Artin, A. Grothendieck and J.L. Verdier (eds), *Théorie des Topos et Cohomologie Étale des Schémas*, SLNM 269, Springer, Berlin (1972).

[A1] I. Aitchison, *The geometry of oriented cubes*, Macquarie Mathematics Reports 86-002 (December 1986).

[A2] I. Aitchison, *String diagrams for non-abelian cocycle conditions*, Handwritten notes, Talk presented at Louvain-la-Neuve, Belgium (1987).

[AS] I. Aitchison and R. Street, *The algebra of oriented cubes*, Handwritten notes.

[ASn] F.A. Al-Agl and R. Steiner, *Nerves of multiple categories*, Proc. London Math. Soc. **66** (1993), 92–128.

[BW] M. Barr and C. Wells, *Toposes, Triples and Theories*, Springer, Berlin (1985).

[Bc] J. Beck, *Distributive laws*, Seminar on Triples and Categorical Homology Theory, SLNM 80, Springer, Berlin (1969), 119–140.

[Bn1] J. Bénabou, *Introduction to bicategories*, Reports of the Midwest Category Seminar, SLNM 47, Springer, Berlin (1967), 1–77.

[Bn2] J. Bénabou, *Les distributeurs*, Seminaires de Math. Pure, Univ. Catholique de Louvain, Rapport No. 33 (1973).

[Bns] D.B. Benson, *The basic algebraic structures in categories of derivations*, Inform. and Control **28** (1975), 1–29.

[BKP] R. Blackwell, G.M. Kelly and A.J. Power, *Two-dimensional monad theory*, J. Pure Appl. Algebra **59** (1989), 1–41.

[BG1] C. Brown and D. Gurr, *Refinement and simulation of nets – a categorical characterisation*, Proc. Application and Theory of Petri Nets, K. Jensen, ed., Lecture Notes in Comput. Sci. vol. 616, Springer, Berlin (1992), 76–92.

[BG2] C. Brown and Doug Gurr, *Timing Petri nets categorically*, Proc. ICALP 1992, W. Kuich, ed., Lecture Notes in Comput. Sci. vol. 623, Springer, Berlin (1992), 571–582.

[BH] R. Brown and P.J. Higgins, *The equivalence of crossed complexes and ∞-groupoids*, Cahiers Topologie Géom. Différentielle Catégoriques **22** (1981), 371–386.

[BS] R. Brown and C.B. Spencer, *G-Groupoids, crossed modules and the fundamental groupoid of a topological group*, Proc. Kon. Nederl. Akad. Wetensch. Ser. A, **79** (1976), 296–302.

[C] D. Conduché, *Modules croisés généralisés de longeur* 2, J. Pure Appl. Algebra **34** (1984), 155–178.

[Da1] B.J. Day, *On closed categories of functors*, Midwest Category Seminar Reports IV, SLNM 137, Springer, Berlin (1970), 1–38.

[Da2] B.J. Day, *A reflection theorem for closed categories*, J. Pure Appl. Algebra **2** (1972), 1–11.

[Da3] B.J. Day, *On closed categories of functors, II*, Category Seminar Sydney 1972/73, SLNM 420, Springer, Berlin (1974), 21–54.

[DM] P. Deligne and J.S. Milne, *Tannakian categories, Hodge cocycles, motives and Shimura varieties*, SLNM 900, Springer, Berlin (1982), 101–228.

[Dr1] V.G. Drinfel'd, *Quantum groups*, Proceedings of the International Congress of Mathematicians at Berkeley, California, USA 1986 (1987), 798–820.

[Dr2] V.G. Drinfel'd, *Quasi-Hopf algebras and Knizhnik–Zamolodchikov equations*, Problems in Modern Quantum Field Theory, A.A. Belavin, A.U. Klimyk and A.B. Zamolodchikov, eds, Research Reports in Physics, Springer, Berlin (1989).

[DS] J. Duskin and R. Street, *Non-abelian cocycles and nerves of n-categories*, In preparation.

[Eh1] C. Ehresmann, *Catégories structurées*, Ann. Sci. École Norm. Sup. **80** (1963), 349–425.

[Eh2] C. Ehresmann, *Catégories et Structures*, Dunod, Paris (1965).

[E] S. Eilenberg, *On normal forms*, Proceedings of the Fourth Biennial Meeting of SEAMS on Modern Applications of Mathematics, S. Nualtarance and Y. Kemprasit, eds, Bangkok (1978), 1–9.

[EK] S. Eilenberg and G.M. Kelly, *Closed categories*, Proc. Conf. Categorical Algebra at La Jolla 1965, Springer, Berlin (1966), 421–562.

[ES] S. Eilenberg and R.H. Street, *Rewrite systems, algebraic structures, and higher-order categories*, Handwritten manuscripts.

[F] P. Freyd, *Abelian Categories*, Harper and Row (1964).

[FY1] P. Freyd and D. Yetter, *Braided compact closed categories with applications to low dimensional topology*, Adv. Math. **77** (1989), 156–182.

[FY2] P. Freyd and D. Yetter, *Coherence theorems via knot theory*, J. Pure Appl. Algebra **78** (1992), 49–76.

[GZ] P. Gabriel and M. Zisman, *Calculus of Fractions and Homotopy Theory*, Springer, Berlin (1967).

[Gd] J. Giraud, *Cohomologie non Abélienne*, Springer, Berlin (1971).

[Gt] R. Godement, *Topologie Algébrique et Théorie des Faisceaux*, Publications de L'Institut de Math de L'Université de Strasbourg XIII, Hermann, Paris (1964).

[GPS] R. Gordon, A.J. Power and R.H. Street, *Coherence for tricategories*, Mem. Amer. Math. Soc., to appear.

[Gy1] J.W. Gray, *Formal Category Theory: Adjointness for 2-Categories*, SLNM 391, Springer, Berlin (1974).

[Gy2] J.W. Gray, *Coherence for the tensor product of 2-categories, and braid groups*, Algebra, Topology, and Category Theory (a collection of papers in honour of Samuel Eilenberg), Academic Press, New York (1976), 63–76.

[H] G. Haet, *Confluent reductions: abstract properties and applications to term rewriting systems*, J. Assoc. Comput. Mach. **27** (1980), 797–821.

[J] M. Johnson, *Pasting Diagrams in n-Categories with Applications to Coherence Theorems and Categories of Paths*, PhD Thesis, University of Sydney, October 1987.

[JW] M. Johnson and R. Walters, *On the nerve of an n-category*, Cahiers Topologie Géom. Différentielle Catégoriques **28** (1987), 257–282.

[Jt] P.T. Johnstone, *Topos Theory*, Academic Press, New York (1978).

[Joy] A. Joyal, *Foncteurs analytiques et espècies de structures*, Combinatoire Énumérative, Proceedings, Montréal, Québec 1985, SLNM 1234, Springer, Berlin (1991), 126–159.

[JS1] A. Joyal and R. Street, *Braided monoidal categories*, Macquarie Math. Reports #850067 (Dec. 1985); Revised #860081 (Nov. 1986).

[JS2] A. Joyal and R. Street, *The geometry of tensor calculus, I*, Adv. Math. **88** (1991), 55–112.

[JS3] A. Joyal and R. Street, *Tortile Yang–Baxter operators in tensor categories*, J. Pure Appl. Algebra **71** (1991), 43–51.

[JS4] A. Joyal and R. Street, *An introduction to Tannaka duality and quantum groups*, Category Theory, Proceedings, Como 1990; Part II of SLNM 1488, Springer, Berlin (1991), 411–492.

[JS5] A. Joyal and R. Street, *Braided tensor categories*, Adv. Math. **102** (1993), 20–78.

[JT] A. Joyal and M. Tierney, *Algebraic homotopy types*, In preparation.

[KV1] M.M. Kapranov and V.A. Voevodsky, *Combinatorial-geometric aspects of polycategory theory: pasting schemes and higher Bruhat orders (List of results)*, Cahiers Topologie Géom. Différentielle Catégoriques **32** (1991), 11–27.

[KV2] M.M. Kapranov and V.A. Voevodsky, ∞-groupoids and homotopy types, Preprint (1990).

[KV3] M.M. Kapranov and V.A. Voevodsky, *2-categories and Zamolodchikov tetrahedra equations*, Proc. Sympos. Pure Math. vol. 56 (1994), 177–259.

[K2] G.M. Kelly, *On MacLane's conditions for coherence of natural associativities, commutativities, etc.*, J. Algebra **1** (1964), 397–402.

[KS] G.M. Kelly and R.H. Street, *Review of the elements of 2-categories*, Category Seminar Sydney 1972/73, SLNM 420, Springer, Berlin (1974), 75–103.

[L] F.W. Lawvere, *Metric spaces, generalized logic, and closed categories*, Rend. Sem. Mat. Fis. Milano **43** (1973), 135–166.

[Lyu] V.V. Lyubashenko, *Tensor categories and RCFT, I. Modular transformations; II. Hopf algebras in rigid categories*, Academy of Sciences of the Ukrainian SSR, Preprints ITP-90-30E and 59E.

[ML1] S. MacLane, *Natural associativity and commutativity*, Rice Univ. Studies **49** (1963), 28–46.

[ML2] S. MacLane, *Categories for the Working Mathematician*, Springer, Berlin (1971).

[MLM] S. MacLane and I. Moerdijk, *Sheaves in Geometry and Logic: A First Introduction to Topos Theory*, Springer, Berlin (1992).

[MP] S. MacLane and R. Paré, *Coherence for bicategories and indexed categories*, J. Pure App. Algebra **37** (1985), 59–80.

[Mj1] S. Majid, *Physics for algebraists: non-commutative and non-cocommutative Hopf algebras by a bicrossproduct construction*, J. Algebra **130** (1990), 17–64.

[MN1] J.M. Maillet and F. Nijhoff, *Multidimensional integrable lattice models, quantum groups, and the d-simplex equations*, Proceedings of the Kiev Conference on Nonliuear and Turbulent Processes in Physics (1989).

[MN2] J.M. Maillet and F. Nijhoff, *The tetrahedron equation and the four-simplex equation*, Phys. Lett. A **134**(4) (2 Jan 1989), 221–228.

[MS] Yu.I. Manin and V.V. Schechtman, *Arrangements of hyperplanes, higher braid groups and higher Bruhat orders*, Advanced Studies in Pure Mathematics vol. 17 (1989), 289–308.

[M] R. Moore, *Apple Macintosh software implementing the excision of extremals algorithm for simplexes and the like*, Macquarie University (1989).

[MS] R. Moore and N. Seiberg, *Classical and quantum conformal field theory*, Comm. Math. Phys. **123**(2) (1989), 177–254.

[PR] R. Penrose and R. Rindler, *Spinors and Space-Time*, Vol. 1, Cambridge Univ. Press, Cambridge, UK (1984).

[Pw1] A.J. Power, *A 2-categorical pasting theorem*, J. Algebra **129** (1990), 439–445.

[Pw2] A.J. Power, *An n-categorical pasting theorem*, Category Theory, Proceedings, Como 1990, SLNM 1488, Springer, Berlin (1991), 326–358.

[P] V. Pratt, *Modeling concurrency with geometry*, Preprint, Stanford University (August 1990).

[RT] N.Yu. Reshetikhin and V.G. Turaev, *Ribbon graphs and their invariants derived from quantum groups*, Comm. Math. Phys. **127**(1) (1990), 1–26.

[SR] N. Saavedra Rivano, *Catégories Tannakiennes*, SLNM 265, Springer, Berlin (1972).

[R1] J.E. Roberts, *Mathematical aspects of local cohomology*, Proc. Colloquium on Operator Algebras and Their Application to Mathematical Physics, Marseille (1977).

[R2] J.E. Roberts, *Complicial sets*, Handwritten notes (1978).

[Sch] H. Schubert, *Categories*, Springer, Berlin (1972).

[Sg] G. Segal, *Classifying spaces and spectral sequences*, Inst. Hautes Études Sci. Publ. Math. **34** (1968), 105–112.

[SMC] Shum Mei Chee, *Tortile Tensor Categories*, PhD Thesis, Macquarie University (November 1989); J. Pure Appl. Algebra **93** (1994), 57–110.

[St] J.D. Stasheff, *Homotopy associativity of H-spaces I, II*, Trans. Amer. Math. Soc. **108** (1963), 275–312.

[Sn] R. Steiner, *The algebra of directed complexes*, Applied Categorical Structures **1** (1993), 247–284.

[Sl] J.G. Stell, *Modelling term rewriting systems by sesqui-categories*, Technical Report TR94-02, Dept. Computer Science, Keele University (January 1994).

[S0] R. Street, *Two constructions on lax functors*, Cahiers Topologie Géom. Différentielle Catégoriques **13** (1972), 217–264.

[S1] R. Street, *The formal theory of monads*, J. Pure Appl. Algebra **2** (1972), 149–168.

[S2] R. Street, *Limits indexed by category-valued 2-functors*, J. Pure Appl. Algebra **8** (1976), 148–181.

[S3] R. Street, *Fibrations in bicategories*, Cahiers Topologie Géom. Différentielle Catégoriques **21** (1980), 111–160; **28** (1987), 53–56.

[S4] R. Street, *Higher-dimensional nerves*, Handwritten notes, April–May 1982.

[S5] R. Street, *The algebra of oriented simplexes*, J. Pure Appl. Algebra **49** (1987), 283–335.

[S6] R. Street, *Parity complexes*, Cahiers Topologie Géom. Différentielle Catégoriques **32** (1991), 315–343.

[S7] R. Street, *Fillers for nerves*, SLNM 1348, Springer, Berlin (1988), 337–341.

[S8] R. Street, *Parity complexes: corrigenda*, Cahiers Topologie Géom. Différentielle Catégoriques **35** (1994), 359–362.

[S9] R. Street, *Higher categories, strings, cubes and simplex equations*, Applied Categorical Structures **3** (1995), 29–77.

[T] V.G. Turaev, *The Yang–Baxter equation and invariants of links*, Invent. Math. **92** (1988), 527–553.

[V] D. Verity, *Higher-dimensional nerves*, Handwritten notes, 1991.

[Y] D.N. Yetter, *Quantum groups and representations of monoidal categories*, Math. Proc. Cambridge Phil. Soc. **108** (1990), 261–290.

[Z] A.B. Zamolodchikov, *Tetrahedra equations and integrable systems in three-dimensional space*, Soviet Phys. JETP **52** (1980), 325; Comm. Math. Phys. **70** (1981), 489.