Journal of Pure and Applied Algebra 19 (1980) 193–213

COHERENCE FOR COMPACT CLOSED CATEGORIES

G.M. KELLY*

Pure Mathematics Department, University of Sydney, N.S.W. 2006, Australia

and

M.L. LAPLAZA

Mathematics Department, University of Puerto Rico, Mayaguez, Puerto Rico 00708

1. Introduction

It gives us great pleasure to honour, on the occasion of his seventieth birthday, our friend and mentor Saunders MacLane; and to acknowledge that our interest in coherence problems stems from his influence, beginning with his fundamental paper [13], and continued through personal contacts—whose value to us, and whose warmth, much transcend what we can express in a formal dedication.

Our purpose is to give an explicit description of the free compact closed category on a given category. A *compact* closed category is a symmetric monoidal one whose internal-hom $[A, C]$ has the form $C \otimes A^*$. Before giving examples of these we analyze and simplify the definition.

A monoidal category $\mathscr{B}$ with tensor product $\otimes$ and unit object I can be regarded as a bicategory $\mathbf{B}$ with a single 0-cell, the 1-cells of $\mathbf{B}$ being the objects of $\mathscr{B}$ with $\otimes$ as their composition, and the 2-cells of $\mathbf{B}$ being the morphisms of $\mathscr{B}$. We can therefore speak of a *left adjoint* of an object A of $\mathscr{B}$, meaning thereby an object A^* of $\mathscr{B}$ together with a "unit" map $d_A : I \to A \otimes A^*$ and a "counit" $e_A : A^* \otimes A \to I$ satisfying the usual "triangular equations", namely that each of the composites

$$A \underset{l^{-1}}{\cong} I \otimes A \xrightarrow[d_A \otimes 1]{} (A \otimes A^*) \otimes A \underset{a}{\cong} A \otimes (A^* \otimes A) \xrightarrow[1 \otimes e_A]{} A \otimes I \underset{r}{\cong} A, \tag{1.1}$$

$$A^* \underset{r^{-1}}{\cong} A^* \otimes I \xrightarrow[1 \otimes d_A]{} A^* \otimes (A \otimes A^*) \underset{a^{-1}}{\cong} (A^* \otimes A) \otimes A^* \xrightarrow[e_A \otimes 1]{} I \otimes A^* \underset{l}{\cong} A^*, \tag{1.2}$$

should be the identity (a, l, and r being the associativity and unit isomorphisms of $\mathscr{B}$).

Whenever A has such a left adjoint we have a natural isomorphism $\mathscr{B}(B \otimes A, C) = \mathscr{B}(B, C \otimes A^*)$; so that if every object in $\mathscr{B}$ has a left adjoint, the monoidal category $\mathscr{B}$ is *closed*, with internal-hom $[A, C] = C \otimes A^*$. Conversely it is easily verified that

* The first author gratefully acknowledges the support of an Australian Research Grant.

an object A in a monoidal closed category has a left adjoint if and only if the canonical map $B\otimes[A,C]\to[A,B\otimes C]$ is an isomorphism for all B and C; for which it suffices that its component $A\otimes[A,I]\to[A,A\otimes I]\cong[A,A]$ be an isomorphism; whereupon A^* is $[A,I]$.

We can therefore adopt the following simple definition of *compact closed category*, which does not explicitly mention any "closed" structure: it is a *symmetric* monoidal category (with symmetry isomorphism $c : A\otimes B\cong B\otimes A$, say), in which *every* object A has a left adjoint; or rather, for our present purposes, one in which every object A has an *assigned* left adjoint (A^*, d_A, e_A)—for left adjoints are unique only to within isomorphism, while we want the extra structure on the category to be equational.

As examples of compact closed categories we have: finitely-generated projective modules over a commutative ring; finite-dimensional representations of a compact group; the category with sets for objects, relations for morphisms, and cartesian product for $\otimes$; Conway's games [2] as made into a category, with strategies as maps, by Joyal (unpublished); Lawvere's (also unpublished) category of physical quantities, in which the objects are physical "dimensions" such as MLT^{-2}. In *any* symmetric monoidal category those objects that *do* have left adjoints form a full subcategory that is compact closed. Those compact closed categories that are ordered sets are precisely the ordered abelian groups; there is clearly a connexion with K-theory. The "category" of small categories and profunctors fails to be a compact closed category only because it fails to be an honest category with associative composition. {The $\star$-*autonomous categories* in the sense of Barr [1] *include* the compact closed categories, but are more general; such a category has an internal-hom of the form $[A,C]=(A\otimes C^*)^*$, and is compact closed exactly when it satisfies $(A\otimes B)^*\cong B^*\otimes A^*$. Having $I^*\cong I$ is not enough to ensure this, as is shown by the example of complete upper semi-lattices}.

Small compact closed categories form a category **Comp** if we take as morphisms those functors that strictly preserve both the symmetric monoidal structure and the assigned adjunctions (A^*, d_A, e_A). The equationality of the extra structure on the underlying category ensures, by general principles, that the forgetful functor $U : \mathbf{Comp}\to\mathbf{Cat}$ has a left adjoint F and is monadic; and in fact a direct proof of this has been given by Day [3], who shows that **Comp** may be obtained from the category of symmetric monoidal closed categories as a category of fractions.

Our goal is to give this left adjoint F (and hence the monad on **Cat**) explicitly, thus completely solving the "coherence problem" for this structure. It is in fact a structure of special interest in the search for a deeper understanding of coherence problems, being perhaps the simplest practical case in which the data and axioms are expressible entirely in terms of the "good" generalized natural transformations of [4], but in which there occur "incompatible" composites, such as

$$I \xrightarrow[d_A]{} A\otimes A^* \xrightarrow[c]{} A^*\otimes A \xrightarrow[e_A]{} I;$$

it was put forward by the first author ([5, p. 102] and [6, p. 134]) as a kind of test case. At the end of the paper we make some remarks on coherence problems in the light of our results.

2. Cycles and traces

For a category $\mathscr{A}$ we define the *set of endomorphisms* to be the disjoint union

$$E(\mathscr{A}) = \sum_{A \in \mathrm{ob}\,\mathscr{A}} \mathscr{A}(A,A),$$

and define the *set of cycles* $[\mathscr{A}]$ to be the quotient set of $E(\mathscr{A})$ modulo the equivalence relation generated by the relation

$$gf \sim fg \quad \text{whenever} \quad f : A \to B \text{ and } g : B \to A.$$

The name "cycle" reflects the fact that, for a "cyclic sequence" of maps

$$A_1 \xrightarrow[f_1]{} A_2 \xrightarrow[f_2]{} A_3 \longrightarrow \cdots \longrightarrow A_{n-1} \xrightarrow[f_{n-1}]{} A_n \xrightarrow[f_n]{} A_1,$$

the different endomorphisms $f_n f_{n-1} \cdots f_2 f_1$ and $f_{n-1} \cdots f_2 f_1 f_n$ have the same image in $[\mathscr{A}]$.

The image of an endomorphism under the canonical projection $\tau : E(\mathscr{A}) \to [\mathscr{A}]$ may be called its *trace*. For any set X, a function $\varrho : E(\mathscr{A}) \to X$ factorizes (uniquely) through τ, as $\varrho = \bar{\varrho}\tau$ say, if and only if $\varrho(gf) = \varrho(fg)$ for $f : A \to B$ and $g : B \to A$; such a ϱ may be called a *trace function*, and τ is the *universal trace function*.

One source of trace-functions is pairs of incompatible natural transformations in the sense of [4]. The general case involves a functor

$$T : \mathscr{A}^{\mathrm{op}} \times \mathscr{A} \times \mathscr{A}^{\mathrm{op}} \times \mathscr{A} \times \cdots \times \mathscr{A}^{\mathrm{op}} \times \mathscr{A} \to \mathscr{B}$$

of $2n$ variables, objects K and L of $\mathscr{B}$, and natural transformations with typical components

$$\alpha : K \to T(A_1, A_1, A_2, A_2, A_3, \ldots, A_{n-1}, A_n, A_n), \tag{2.1}$$

$$\beta : T(B_1, B_2, B_2, B_3, B_3, \ldots, B_n, B_n, B_1) \to L; \tag{2.2}$$

and the result is (cf. [5, p. 102]):

Proposition 2.1. *For maps* $f_1 : B_1 \to A_1$, $f_2 : A_1 \to B_2$, $f_3 : B_2 \to A_2$,

$$f_4 : A_2 \to B_3, \ldots, f_{2n-1} : B_n \to A_n,\ f_{2n} : A_n \to B_1,$$

the composite of the three maps (2.1), $T(f_1, f_2, \ldots, f_{2n-1}, f_{2n})$, *and* (2.2), *depends only on the cycle* $\tau(f_{2n} f_{2n-1} \ldots f_2 f_1)$*; so that* α *and* β *give rise to a function* $\gamma : [\mathscr{A}] \to \mathscr{B}(K, L)$.

Proof. For simplicity we give the proof only for $n=1$, this being the only case we actually use below. The above composite then takes the form

$$K \xrightarrow[\alpha_A]{} T(A,A) \xrightarrow[T(g,f)]{} T(B,B) \xrightarrow[\beta_B]{} L,$$

which we may denote by $\delta(g,f)$. The naturality of α gives $T(g,1_A)\alpha_A = T(1_B,g)\alpha_B$, whence $\delta(g,f)=\beta_B T(1_B,f)T(1_B,g)\alpha_B = \delta(1_B,fg)$ depends only on the endomorphism fg. But the naturality of β gives $\beta_B T(g,1_B)=\beta_A T(1_A,g)$, so that we also have $\delta(g,f)=\delta(1_A,gf)$. □

Note that any functor $T:\mathscr{A}\to\mathscr{B}$ sends endomorphisms of $\mathscr{A}$ to endomorphisms of $\mathscr{B}$, and passes to the quotient to give a function $[T]:[\mathscr{A}]\to[\mathscr{B}]$.

The remaining remarks about cycles and traces are not used below, and we include them for general interest. It is easily seen that the function $[T]$ above depends only on the isomorphism-class of T. Since clearly $[\mathscr{A}\times\mathscr{B}]=[\mathscr{A}]\times[\mathscr{B}]$, a functor of the form $\otimes:\mathscr{A}\times\mathscr{A}\to\mathscr{A}$ induces a binary operation $[\otimes]$ on $[\mathscr{A}]$, and similarly for ternary functors $\mathscr{A}^3\to\mathscr{A}$ and so on. Moreover $[\otimes]$ is associative [commutative] if $\otimes$ is associative [commutative] to within isomorphism. Hence if $\mathscr{A}$ is [symmetric] monoidal, $[\mathscr{A}]$ is a [commutative] monoid. Then if X is a monoid, $\bar{\varrho}:[\mathscr{A}]\to X$ is a monoid-map if and only if $\varrho=\bar{\varrho}\tau$ satisfies $\varrho(k\otimes k')=\varrho(k)\varrho(k')$ and $\varrho(1_I)=1$; in particular τ satisfies these. Further, since $[\mathscr{A}^{\mathrm{op}}]=[\mathscr{A}]$, an internal-hom $\mathscr{A}^{\mathrm{op}}\times\mathscr{A}\to\mathscr{A}$ gives another binary operation on $[\mathscr{A}]$, and so on.

It is immediate that $\varrho:E(\mathscr{A})\to X$ is a trace function if and only if its components $\varrho_A:\mathscr{A}(A,A)\to X$ constitute a natural transformation in the sense of [4]; whence $[\mathscr{A}]$ is the coend $\int^A \mathscr{A}(A,A)$. If now $\mathscr{A}$ is replaced by a $\mathscr{V}$-category for a symmetric monoidal closed category $\mathscr{V}$, we can define $[\mathscr{A}]\in\mathscr{V}$ by this same formula, interpreted now as a $\mathscr{V}$-coend; and everything above carries over, with $\times$ replaced by $\otimes$. Note that for an additive category $\mathscr{A}$, this **Ab**-based $[\mathscr{A}]\in\mathbf{Ab}$ is quite different from our original $[\mathscr{A}]\in\mathbf{Set}$ for the underlying category; the former is an *abelian group* which is universal for *additive* trace functions—those for which $\varrho(k+k')=\varrho(k)+\varrho(k')$. The reader may find it instructive to compute $[\mathscr{A}]$ when $\mathscr{A}$ is the (ordinary) category of finite sets, along with the operations on $[\mathscr{A}]$ induced by coproduct, by product, and by exponentiation; and to compute the R-linear $[\mathscr{A}]$ when $\mathscr{A}$ is the category of matrices over a commutative ring R. We observe finally that $[T]:[\mathscr{A}]\to[\mathscr{B}]$ can be defined not only for a $\mathscr{V}$-functor T but for an adjoint pair of $\mathscr{V}$-profunctors; so that $[\mathscr{A}]$ really depends only on the Cauchy completion of $\mathscr{A}$ in the sense of Lawvere [12] (who also has some brief remarks about traces in his Section 5).

3. The explicit description of $F\mathscr{A}$

We define a *signed set* P to be a set $|P|$ together with a "sign" function from P to the two-element set $\{-,+\}$. We write $P\otimes Q$ for the disjoint union of signed sets, with I for the empty signed set; we write P^* for P with the signs reversed; and we write 1 for the one-element signed set with sign $+$.

By an *involution* θ we mean a category which is a coproduct of copies of the arrow-category **2**; these are really the *fixed-point-free* involutions, but we need no others. Such an involution is a special kind of *order* on its object-set $|P|$; and, as with orders in general, we do not distinguish two involutions on $|P|$ which differ only in the "names" of their maps. This object-set $|P|$ becomes *signed* when we attribute $-$ to the source and $+$ to the target of each arrow **2**; we call θ *an involution on the signed set* P when this signing agrees with that of P.

By a *loop* L we mean the free category on a graph of the form

$$A_1 \to A_2 \to \cdots \to A_n \to A_1,$$

for some $n \geq 1$. The composite represented by this string determines a canonical element $\langle L \rangle$ of the set $[L]$ of cycles of L, independently of where the loop is started.

We use $+$ for the coproduct of categories, and henceforth suppose all signed sets and involutions to be *finite*. If θ is an involution on the signed set $P^* \otimes Q$ and ϕ is one on $Q^* \otimes R$, consider the pushout $\theta +_{|Q|} \phi$ in **Cat**, obtained from the coproduct $\theta + \phi$ by identifying the two copies of the discrete category $|Q|$, one in θ and one in ϕ. Write $\phi\theta$ for the full subcategory of $\theta +_{|Q|} \phi$ determined by the object-set $|P| + |R|$; it is clearly an involution on $P^* \otimes R$; moreover we clearly have $\theta +_{|Q|} \phi = \phi\theta + \phi \star \theta$, where $\phi \star \theta$ is a coproduct $\sum L_i$ of loops with objects in $|Q|$. If now ψ is a third involution on $R^* \otimes S$, it is further immediate that

$$\psi(\phi\theta) = (\psi\phi)\theta \quad \text{and} \quad \psi \star (\phi\theta) + \phi \star \theta = (\psi\phi) \star \theta + \psi \star \phi. \tag{3.1}$$

For any category $\mathscr{A}$ we now define an explicit compact closed category $G\mathscr{A}$ and a functor $\Psi\mathscr{A} : \mathscr{A} \to G\mathscr{A}$. Our main theorem below in Section 8 will be a proof that $G\mathscr{A}$ is (isomorphic to) the free compact closed category $F\mathscr{A}$ on $\mathscr{A}$, with $\Psi\mathscr{A}$ as the unit of the adjunction. Until then, we need the distinct names $F\mathscr{A}$ and $G\mathscr{A}$ for clarity.

We define the objects of $G\mathscr{A}$ as words in an abstract language:

each object of $\mathscr{A}$ is an object of $G\mathscr{A}$;
I is an object of $G\mathscr{A}$;
for any objects X, Y of $G\mathscr{A}$ there is an object $X \otimes Y$ of $G\mathscr{A}$;
for any object X of $G\mathscr{A}$ there is an object X^* of $G\mathscr{A}$.

To each such object X we assign inductively a signed set $P(X)$, setting $P(A) = 1$ if $A \in \mathscr{A}$, $P(I) = I$, $P(X \otimes Y) = P(X) \otimes P(Y)$, and $P(X^*) = (P(X))^*$. To each such X we also assign inductively an "argument" function $\alpha_X : |P(X)| \to \mathrm{ob}\,\mathscr{A}$; here α_A for $A \in \mathscr{A}$ sends 1 to A; α_I is the unique function from the empty set $|I|$; $\alpha_{X \otimes Y}$ is the function (α_X, α_Y) from the disjoint union; and $\alpha_{X^*} = \alpha_X$.

For any set V we denote by MV the free commutative monoid on V, written additively.

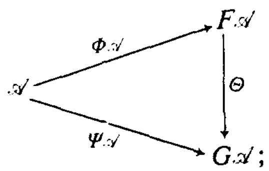
We define a morphism $X \to Y$ in $G\mathscr{A}$ to be a triple (θ, p, λ), where θ is an involution on $(P(X))^* \otimes P(Y) = P(X^* \otimes Y)$, where p is a functor $p : \theta \to \mathscr{A}$ whose value on objects is given by $\alpha_{X^* \otimes Y}$, and where λ is an element of $M[\mathscr{A}]$.

Given another morphism $(\phi, q, \mu) : Y \to Z$, write s and t respectively for the restrictions of the functor $(p,q) : \theta +_{|P(Y)|} \phi \to \mathscr{A}$ to the subcategories $\phi\theta$ and $\phi \star \theta$ ($= \sum L_i$, say). Then we define the composite of (θ, p, λ) and (ϕ, q, μ) to be $(\phi\theta, s, \lambda + \mu + \sum_i [t]\langle L_i \rangle)$. This composition is associative by (3.1); and the identity $(\theta, p, \lambda) : X \to X$ has for θ the obvious involution on $(P(X))^* \otimes P(X)$, while p takes every map to an identity, and $\lambda = 0$. Some pictorial representations of composition-laws of this general kind are given in [5, pp. 98–101] and [6, p. 130].

The tensor product of $G\mathscr{A}$ is given on objects by the formal $\otimes$ of words, and on maps by $(\theta, p, \lambda) \otimes (\phi, q, \mu) = (\theta + \phi, (p,q), \lambda + \mu)$; it is clearly a functor. The operation $(\ )^*$ is needed only on objects, and is the formal $(\ )^*$. Each component of the basic data a, r, c, d, e is of the form (θ, p, λ) where θ is the obvious involution, p sends every map to an identity, and $\lambda = 0$. Clearly a, r, c are natural isomorphisms satisfying the coherence conditions of [13] (l is not needed, being defined in terms of r and c); while d and e satisfy (1.1) and (1.2). Thus $G\mathscr{A}$ is a compact closed category.

Finally we define $\Psi\mathscr{A} : \mathscr{A} \to G\mathscr{A}$ to be the functor sending A to A and sending $f : A \to B$ to $(\mathbf{2}, f, 0)$, where $\mathbf{2}$ is the unique involution on $P(A^* \otimes B) = 1^* \otimes 1$ and f loosely denotes the functor $\mathbf{2} \to \mathscr{A}$ sending the non-identity map of $\mathbf{2}$ to $f : A \to B$.

If $F\mathscr{A}$ denotes the free compact closed category on $\mathscr{A}$, with unit $\Phi\mathscr{A} : \mathscr{A} \to F\mathscr{A}$, we have a unique functor Θ in **Comp** rendering commutative



(3.2)

and ultimately we shall prove Θ to be an isomorphism. But even without that knowledge there is one conclusion we can draw already: since $\Psi\mathscr{A}$ is clearly faithful, it follows from (3.2) that

Proposition 3.1. *The adjunction-unit* $\Phi\mathscr{A} : \mathscr{A} \to F\mathscr{A}$ *is faithful.*

4. Generators and relations for $F\mathscr{A}$

Starting with a given category $\mathscr{A}$ we now describe by generators and relations a compact closed category $F\mathscr{A}$ that is clearly the free one on $\mathscr{A}$. The method is perfectly general, applying to categories with *any* explicitly-given equational extra structure. It proves existence, and allows us to make some general assertions about the nature of $F\mathscr{A}$. To pass from this generators-and-relations description of $F\mathscr{A}$ to an explicit description in closed form, however, involves the solution of a word-problem: which may be called the *coherence problem* for the structure in question.

For compact closed categories, since none of the axioms involves an equation between objects, it is easy to give the objects of $F\mathscr{A}$: they are just what we called in § 3 the objects of $G\mathscr{A}$, which form the *free* $(\otimes, I, (\ )^*)$*-algebra* on ob $\mathscr{A}$.

We next describe the arrows of a graph $H\mathscr{A}$ with these objects. For objects X, Y, Z of $F\mathscr{A}$ there are to be arrows

$$a_{XYZ} : (X\otimes Y)\otimes Z \to X\otimes(Y\otimes Z), \qquad \bar{a}_{XYZ} : X\otimes(Y\otimes Z)\to(X\otimes Y)\otimes Z,$$

$$r_X : X\otimes I\to X, \qquad \bar{r}_X : X\to X\otimes I, \qquad c_{XY} : X\otimes Y\to Y\otimes X,$$

$$d_X : I\to X\otimes X^*, \quad \text{and} \quad e_X : X^*\otimes X\to I;$$

while for each $f : A\to B$ in $\mathscr{A}$ there is to be an arrow $\{f\} : A\to B$ in $H\mathscr{A}$. These arrows may be called *formal instances* of $a, \ldots, f$. We complete the description of the arrows of $H\mathscr{A}$ by decreeing that, whenever $t : X\to Y$ is such an arrow and Z is an object of $F\mathscr{A}$, there are to be arrows $Z\otimes t : Z\otimes X\to Z\otimes Y$ and $t\otimes Z : X\otimes Z\to Y\otimes Z$. All of these arrows live of course in a formal language, and are distinct if they have different names. If we define an *expansion* of t to be an arrow of one of the forms t, $Z\otimes t$, $t\otimes Z$, $W\otimes(Z\otimes t)$, $W\otimes(t\otimes Z)$, and so on, then the arrows of $H\mathscr{A}$ are precisely the expansions of the formal instances of $a, \ldots, f$; and may be called the *formal expanded instances*.

We now describe a relation $\mathfrak{r}\mathscr{A}$, or $\sim$ for short, on the free category $K\mathscr{A}$ generated by the graph $H\mathscr{A}$. As is usual in word problems, we also use the word "relation" for an *element* of $\mathfrak{r}\mathscr{A}$. In this language, the relations that make up $\mathfrak{r}\mathscr{A}$ are as follows. First, there are the relations

$$(t\otimes W)(X\otimes s)\sim(Y\otimes s)(t\otimes Z) : X\otimes Z\to Y\otimes W$$

and

$$1_X\otimes Y\sim X\otimes 1_Y\sim 1_{X\otimes Y},$$

that we need for $\otimes$ to be a functor. Then there are the relations that assert the naturality of $a, \bar{a}, r, \bar{r}, c$. Next there are the coherence relations of [13] for a, r, and c, along with $a\bar{a}\sim 1$, $\bar{a}a\sim 1$, $r\bar{r}\sim 1$, and $\bar{r}r\sim 1$. Then come the relations (1.1) and (1.2) for d and e. And then the relation $\{f\}\{g\}=\{fg\}$ for each composable pair of maps in $\mathscr{A}$, along with $1_A=\{1_A\}$, that we need for $\mathscr{A}\to F\mathscr{A}$ to be a functor. Finally, we need all the *expansions* of these relations.

We end by defining $F\mathscr{A}$ as the quotient of the free category $K\mathscr{A}$ on $H\mathscr{A}$ modulo the relation $\mathfrak{r}\mathscr{A}$; that is, modulo the category-congruence $\mathfrak{q}\mathscr{A}$ generated by $\mathfrak{r}\mathscr{A}$. We recall that $g, h : X\to Y$ in $K\mathscr{A}$ are related under $\mathfrak{q}\mathscr{A}$, or equivalently have the same image in $F\mathscr{A}$, exactly where there is a sequence $g=k_0, k_1, \ldots, k_n=h$ in which each k_{i-1}, k_i has the form ypx, yqx, where p and q are related, in one order or the other, under $\mathfrak{r}\mathscr{A}$. Thus finding the maps $X\to Y$ in $F\mathscr{A}$ is exactly a *word problem* in the classical sense of Thue, except for the fact that the "letters" (the arrows of $H\mathscr{A}$) are not always composable. In another language, we may express the relation of $\mathfrak{q}\mathscr{A}$ to $\mathfrak{r}\mathscr{A}$ by saying that a *diagram* g, h in $K\mathscr{A}$ commutes in $F\mathscr{A}$ if and only if it can be "filled in" with diagrams p, q from $\mathrm{r}\mathscr{A}$; which makes it reasonable to call the determination of $\mathrm{q}\mathscr{A}$ the *coherence problem*.

It is clear from the construction that $F\mathscr{A}$ does admit a canonical structure of compact closed category, and that it is the free one on $\mathscr{A}$, the unit $\Phi\mathscr{A}: \mathscr{A} \to F\mathscr{A}$ being the functor sending A to A and f to $\{f\}$. By Proposition 3.1 it can do no harm to replace $\{f\}$ by the simpler f, and we henceforth do so.

The one general assertion about the nature of $F\mathscr{A}$, apart from the precise description of its objects, that follows from the above construction, and that we need for the sequel, is the following. If $\mathscr{B}$ is any compact closed category, we can define an *instance* in $\mathscr{B}$ of any one of $a, \dots, e$ to be an actual component such as a_{ABC} or d_A; and if we have some functor $\mathscr{A} \to \mathscr{B}$ we can define an *instance* in $\mathscr{B}$ of a map f in $\mathscr{A}$ as its image. We can define an *expanded instance* in $\mathscr{B}$ of one of $a, \dots, f$ to be the result of starting with an actual instance and tensoring repeatedly with identity maps. Since the images in $F\mathscr{A}$ of formal instances and formal expanded instances in $H\mathscr{A}$ are clearly actual instances and expanded instances, we have:

Proposition 4.1. *Every map in $F\mathscr{A}$ is a composite of expanded instances of a, a^{-1}, r, r^{-1}, c, d, e, and maps f in $\mathscr{A}$.* □

5. Central maps

To handle expeditiously that fragment of the word problem represented by the *central* maps—those that involve only the symmetric-monoidal data a, a^{-1}, r, r^{-1}, c—we use a suitable form of MacLane's coherence theorem [13].

We define a category $\mathscr{P}$ whose objects form the free $(\otimes, I)$-algebra on a single "formal variable" **1**, assigning to each such object a numerical "arity" ΓT, namely the number of **1**'s in it. The maps $T \to S$ in $\mathscr{P}$ are to be the bijections $\Gamma T \to \Gamma S$; there are none unless $\Gamma T = \Gamma S = n$ say, and in that case they are the permutations ξ of n. The composition is that of permutations.

For any category $\mathscr{A}$ we define a category $\mathscr{P} \circ \mathscr{A}$. An object is an expression $T[A_1, \dots, A_n]$ where $T \in \mathscr{P}$ with $\Gamma T = n$ and where the $A_i \in \mathscr{A}$. The maps have the form $\xi[f_1, \dots, f_n]: T[A_1, \dots, A_n] \to S[B_1, \dots, B_n]$ where $\xi: T \to S$ in $\mathscr{P}$ and where the $f_i: A_{\xi^{-1}i} \to B_i$ are maps in $\mathscr{A}$; composition is evident. We use the convention that the name of an object is the name of its identity map, in writing such special maps as

$$T[f_1, \dots, f_n]: T[A_1, \dots, A_n] \to T[B_1, \dots, B_n]$$

where $f_i: A_i \to B_i$, or

$$\xi[A_1, \dots, A_n]: T[A_{\xi 1}, \dots, A_{\xi n}] \to S[A_1, \dots, A_n].$$

As was shown in [6] (where $\mathscr{P}$ was called the *club* for symmetric monoidal categories), MacLane's coherence theorem is equivalent to the assertion that $\mathscr{P} \circ \mathscr{A}$ is the free symmetric monoidal category on $\mathscr{A}$. Thus for every symmetric monoidal category $\mathscr{B}$, the structure is given by an *action* $\mathscr{P} \circ \mathscr{B} \to \mathscr{B}$. We write $T(A_1, \ldots, A_n)$ and $\xi(f_1, \ldots, f_n)$ for the images of $T[A_1, \ldots, A_n]$ and $\xi[f_1, \ldots, f_n]$ under this action; thus for instance $((1 \otimes 1) \otimes I)(A, B)$ denotes $(A \otimes B) \otimes I$.

In particular we can apply this language when $\mathscr{B}$ is one of the compact closed categories $F\mathscr{A}$ or $G\mathscr{A}$. First, define an object of $F\mathscr{A}$ to be *prime* if it is either an object A of $\mathscr{A}$ or else of the form Y^*. Then every object of $F\mathscr{A}$ clearly has a *unique* expression as $T(X_1, \ldots, X_n)$ with $T \in \mathscr{P}$ and with the X_i prime; we call this its *prime factorization*. Observe that n may be 0: the prime factorization of I is $I(\ )$.

We now define a *central map* $x : Y \to Z$, whether in $F\mathscr{A}$ or in $G\mathscr{A}$, as one that can be written in the form $\xi(X_1, \ldots, X_n) : T(X_{\xi 1}, \ldots, X_{\xi n}) \to S(X_1, \ldots, X_n)$ *with* $T, S \in \mathscr{P}$ and the X_i prime. In the case of $G\mathscr{A}$ this map x can be identified at once as $(\theta, p, 0)$, where θ is the evident involution on $P(Y^* \otimes Z)$ corresponding to the permutation ξ, and p sends every map to an identity. If none of the $P(X_i)$ is empty—that is, if no X_i is a *constant* such as $(I \otimes I^*)^*$—then ξ is determined in turn by θ, and thus by the map x of $G\mathscr{A}$. Since the strict functor Θ of (3.2) sends the central map $\xi(X_1, \ldots, X_n)$ of $F\mathscr{A}$ to the map also so denoted in $G\mathscr{A}$, it follows that in $F\mathscr{A}$ too a central map $x : Y \to Z$ can be written as $\xi(X_1, \ldots, X_n)$ in only *one* way, provided that none of the primes X_i is constant. We can then speak of ξ as *the* association of the prime factors of Y to those of Z corresponding to the central map x.

The following properties of the central maps in $F\mathscr{A}$ or in $G\mathscr{A}$ are evident (cf. [7, p. 200]): $\xi(X_1, \ldots, X_n)$ is central even if the X_i are not prime; $T(x_1, \ldots, x_n)$ is central if each x_i is; they constitute a symmetric monoidal subcategory of $F\mathscr{A}$ or $G\mathscr{A}$, consisting entirely of isomorphisms; and they coincide with the composites of the expanded instances of the central data a, a^{-1}, r, r^{-1}, c.

6. Elementary properties of compact closed categories

We next describe certain "derived operations" of the theory of compact closed categories, which will be used later to simplify the description of $F\mathscr{A}$.

The "pasting" composition of diagrams containing 2-cells works as smoothly in a bicategory as in a 2-category, provided that the appropriate associativity isomorphisms, connecting the various composites of 1-cells, are incorporated into the *definition* of "pasting". It follows that the simple discussion in [11, § 2] of adjunction in a 2-category applies with trivial modifications to a bicategory, providing immediate proofs of the assertions below. A reader preferring more direct proofs will find it easy, if somewhat tedious, to give them. We consider throughout a compact closed category $\mathscr{B}$.

First, there is by Proposition 2.1 of [11] a bijection between maps $f : A \to B$ in $\mathscr{B}$ and maps $f^* : B^* \to A^*$ between their left adjoints. The point of that proposition is that the triangular equations for the unit and the counit of an adjunction—in our case the equations (1.1) and (1.2) for d_A and e_A—express them as two-sided *pasting* inverses. Thus the bijection between f and f^* can be determined by any one of four *equivalent* equations, whose respective analogues in ordinary associative algebra have the forms $d'f^*=fd$, $f^*e=e'f$, $f^*=e'fd$, and $d'f^*e=f$, in which d, e and d', e' are supposed to be pairs of inverses.

In the present setting one of the four equations asserts the equality of the two maps $(f\otimes 1)d_A$ and $(1\otimes f^*)d_B$ from I to $B\otimes A^*$, and another the equality of the maps $e_B(1\otimes f)$ and $e_A(f^*\otimes 1)$ from $B^*\otimes A$ to I: in other words the *naturality* in the sense of [4] of d and e. A third gives f^* explicitly in terms of f, as

$$\begin{array}{ccc} B^* \overset{r^{-1}}{\cong} B^*\otimes I & \xrightarrow{1\otimes d_A} & B^*\otimes(A\otimes A^*) \\ {\scriptstyle f^*}\downarrow & & \downarrow{\scriptstyle 1\otimes(f\otimes 1)} \\ A^* \underset{l}{\cong} I\otimes A^* & \xleftarrow[e_B\otimes 1]{} (B^*\otimes B)\otimes A^* \underset{a^{-1}}{\cong} & B^*\otimes(B\otimes A^*); \end{array} \tag{6.1}$$

and the fourth, which we do not need, gives f in terms of f^*.

That this definition of f^* actually makes $(\ )^*$ into a functor $\mathscr{B}^{\mathrm{op}}\to\mathscr{B}$ follows from Proposition 2.2 of [11]. Thus in fact there is exactly one way of making $(\ )^*$ functorial that makes d and e natural; and we could have included this functoriality and naturality as part of the definition of compact closed category. However for our purposes this would complicate matters unnecessarily, by increasing both the data to be given and the axioms to be verified.

Since adjunctions compose, *one* left adjoint of $A\otimes B$ is $B^*\otimes A^*$, with the appropriate d and e. Hence, by Proposition 2.3 of [11], this is related to the *assigned* left adjoint $(A\otimes B)^*$ by a canonical isomorphism $u_{AB}:(A\otimes B)^*\cong B^*\otimes A^*$. There are now eight equivalent equations each serving to determine u: four for u and four for u^{-1}. One of these gives u explicitly as a composite of expanded instances of the basic data $a, a^{-1}, r, r^{-1}, c, d, e$; and another does the same for u^{-1}. One of the others gives $d_{A\otimes B}$ in terms of d_A and d_B as

$$\begin{array}{ccc} I & \xrightarrow{d_A} & A\otimes A^*\cong(A\otimes I)\otimes A^* \\ {\scriptstyle d_{A\otimes B}}\downarrow & & \downarrow{\scriptstyle (1\otimes d_B)\otimes 1} \\ (A\otimes B)\otimes(A\otimes B)^* & \xrightarrow[1\otimes u_{AB}]{\cong} & (A\otimes B)\otimes(B^*\otimes A^*)\cong(A\otimes(B\otimes B^*))\otimes A^*; \end{array} \tag{6.2}$$

while another does the analogue for $e_{A\otimes B}$.

Similarly, since I is a left adjoint of itself, we have an isomorphism $v: I^*\cong I$, one of the defining equations for which gives d_I as

$$\begin{array}{ccc} I & & \\ {\scriptstyle d_I}\downarrow & \searrow{\scriptstyle \cong} & \\ I\otimes I^* & \xrightarrow[1\otimes v]{\cong} & I\otimes I, \end{array} \tag{6.3}$$

while another gives the analogue for e_I.

Again, because $\mathscr{B}$ is symmetric, A is one left adjoint of A^*, and we have an isomorphism $w_A : A^{**} \cong A$, one of the defining equations for which gives d_{A^*} as

$$\begin{array}{ccc} I & \xrightarrow{d_A} & A \otimes A^* \\ {\scriptstyle d_{A^*}} \downarrow & & \downarrow {\scriptstyle c}\ \cong \\ A^* \otimes A^{**} & \xrightarrow[1\otimes w_A]{\cong} & A^* \otimes A, \end{array} \tag{6.4}$$

while another gives the analogue for e_{A^*}.

We next need:

Proposition 6.1. *In any monoidal category $\mathscr{B}$ (symmetric or not), the endomorphism-monoid $\mathscr{B}(I,I)$ is commutative, and moreover the value of the composite*

$$I \cong I \otimes I \xrightarrow[f\otimes g]{} I \otimes I \cong I$$

is $fg = gf$.

Proof. By naturality, the composites

$$I \xrightarrow[r_I^{-1}]{} I \otimes I \xrightarrow[f\otimes 1]{} I \otimes I \xrightarrow[r_I]{} I, \qquad I \xrightarrow[l_I^{-1}]{} I \otimes I \xrightarrow[1\otimes g]{} I \otimes I \xrightarrow[l_I]{} I,$$

are f and g respectively; but the isomorphisms l_I and r_I coincide by coherence, whence $fg = gf$ since $(f\otimes 1)(1\otimes g) = (1\otimes g)(f \otimes 1) = f \otimes g$. □

Returning to a compact closed category $\mathscr{B}$, define $\delta(g,f)$ for $f : A \to B$ and $g : B \to A$ as the composite

$$I \xrightarrow[d_A]{} A \otimes A^* \underset{c}{\cong} A^* \otimes A \xrightarrow[g^* \otimes f]{} B^* \otimes B \xrightarrow[e_B]{} I. \tag{6.5}$$

The part of the following that refers to the multiplicative structure of $[\mathscr{B}]$ (see the penultimate paragraph of Section 2) is not actually needed below:

Proposition 6.2. *$\delta(g,f)$ depends only on the cycle $\tau(gf) = \tau(fg)$, and therefore defines a function $\gamma : [\mathscr{B}] \to \mathscr{B}(I,I)$. In fact γ is a map of commutative monoids with involution, satisfying $\gamma(k \otimes k') = \gamma(k)\gamma(k')$, $\gamma(1_I) = 1$, $\gamma(k^*) = \gamma(k)$.*

Proof. The first assertion follows from Proposition 2.1, since d, e and c are natural. For the second we merely indicate the proof: first observe that γ does not change if we alter the choice of adjoints, and then observe that the above equations are immediate if u, v, and w are identities. □

7. Reduced maps in $F\mathscr{A}$

As the next step in solving the word problem, we show in effect that those expanded instances of d_X and e_X for which X is not a mere object of $\mathscr{A}$ can be absorbed inside isomorphisms at the ends of the word.

Let us agree that in future "expanded instance" means "expanded instance of one of $a, a^{-1}, r, r^{-1}, c, d, e$ or of some map f in $\mathscr{A}$". Now define a *reduced* object of $F\mathscr{A}$ to be one whose prime factors are all of the form A or A^* for $A \in \mathscr{A}$; and call a map $X \to Y$ in $F\mathscr{A}$ *reduced* if it can be written for some $n \geq 0$ in the form

$$X = Z_0 \xrightarrow{t_1} Z_1 \xrightarrow{t_2} Z_2 \longrightarrow \cdots \xrightarrow{t_n} Z_n = Y, \tag{7.1}$$

where each t_i is an expanded instance and each Z_i is reduced: which implies that X and Y are themselves reduced. Clearly a t_i in (7.1) cannot then be an expanded instance of d_V or of e_V unless V is an object A of $\mathscr{A}$. The purpose of this section is now to prove:

Proposition 7.1. *Every object in $F\mathscr{A}$ is isomorphic to a reduced one, and every map in $F\mathscr{A}$ between reduced objects is reduced.*

We begin by assigning to each object Z of $F\mathscr{A}$ a numerical *rank* $\kappa(Z)$. We first define it inductively for *prime* Z by

$$\begin{aligned}
&\kappa(A) = 0,\\
&\kappa(A^*) = 0,\\
&\kappa(I^*) = 1,\\
&\kappa((X \otimes Y)^*) = \kappa(X^*) + \kappa(Y^*) + 1,\\
&\kappa(X^{**}) = 3\kappa(X^*) + 1,
\end{aligned}$$

where A denotes an object of $\mathscr{A}$; then for any Z, with prime factorization $T(X_1, \ldots, X_n)$, we set $\kappa(Z) = \sum \kappa(X_i)$. Clearly $\kappa(T(X_1, \ldots, X_n))$ is then $\sum \kappa(X_i)$ whether the X_i are prime or not; or equivalently $\kappa(X \otimes Y) = \kappa(X) + \kappa(Y)$ and $\kappa(I) = 0$. It is immediate that $\kappa(Z) = 0$ if and only if Z is reduced.

We next assign to each object Z of $F\mathscr{A}$ another object RZ and an isomorphism $\varrho Z : Z \to RZ$ in $F\mathscr{A}$. If Z has prime factorization $T(X_1, \ldots, X_n)$ we set $RZ = T(RX_1, \ldots, RX_n)$ and $\varrho Z = T(\varrho X_1, \ldots, \varrho X_n)$; so it suffices then to give the definition for prime Z. For the reduced primes A and A^* with $A \in \mathscr{A}$, we set $RA = A$, $\varrho A = 1$, $RA^* = A^*$, $\varrho A^* = 1$. For the remaining primes of the form $Z = V^*$, we use in the case of reduced V the isomorphisms u, v, w of (6.2)–(6.4) to define

$$\begin{aligned}
&R((X \otimes Y)^*) = Y^* \otimes X^*, && \varrho((X \otimes Y)^*) = u_{XY}, && \text{for } X, Y \text{ reduced};\\
&R(I^*) = I, && \varrho(I^*) = v;\\
&R(A^{**}) = A, && \varrho(A^{**}) = w_A, && \text{for } A \in \mathscr{A}.
\end{aligned}$$

For the primes V^* with V not reduced, we complete the definition inductively, using the functorial character of $(\ )^*$ from Section 6:

$$R(V^*)=(RV)^*, \qquad \varrho(V^*)=((\varrho V)^*)^{-1}, \quad \text{for } V \text{ not reduced.}$$

Clearly for reduced Z we have $RZ=Z$ and $\varrho Z=1$. The first assertion of Proposition 7.1 follows immediately from

Lemma 7.2. *If Z is not reduced, $\kappa(RZ)<\kappa(Z)$.*

Proof. It suffices to check it for prime Z, and it is immediate if $Z=V^*$ with V reduced. So it remains to prove that

$$\kappa((RV)^*)<\kappa(V^*) \quad \text{for } V \text{ not reduced.} \tag{7.2}$$

We prove this by induction (on the complexity of V). If $V=X\otimes Y$, the respective ranks of $(RV)^*=(RX\otimes RY)^*$ and of $V^*=(X\otimes Y)^*$ are $\kappa((RX)^*)+\kappa((RY)^*)+1$ and $\kappa(X^*)+\kappa(Y^*)+1$, giving (7.2) by induction since X and Y are not both reduced. If $V=W^*$ with W not reduced, the respective ranks of $(RV)^*=(RW)^{**}$ and of $V^*=W^{**}$ are $3\kappa((RW)^*)+1$ and $3\kappa(W^*)+1$, again giving (7.2) by induction. There remain the cases $V=W^*$ with W reduced. If V is I^* or A^{**} for $A\in\mathcal{A}$, (7.2) is immediate since $\kappa(I^*)<\kappa(I^{**})$ and $\kappa(A^*)<\kappa(A^{***})$. Finally if $V=(X\otimes Y)^*$ with X and Y reduced, (7.2) follows from the calculations

$$\kappa((Y^*\otimes X^*)^*)=3\kappa(X^*)+3\kappa(Y^*)+3$$

and

$$\kappa((X\otimes Y)^{**})=3\kappa(X^*)+3\kappa(Y^*)+4. \qquad \square$$

To complete the proof of Proposition 7.1 we use:

Lemma 7.3. *For any expanded instance $t: X\to Y$ in $F\mathcal{A}$, there is a commutative diagram*

$$\begin{array}{ccccccccc}
X & & & \xrightarrow{\quad t \quad} & & & & & Y \\
\downarrow{\scriptstyle \varrho X} & & & & & & & & \downarrow{\scriptstyle \varrho Y} \\
RX=V_0 & \xrightarrow[s_1]{} & V_1 & \xrightarrow[s_2]{} & V_2 & \longrightarrow & \cdots & \xrightarrow[s_n]{} & V_n=RY
\end{array} \tag{7.3}$$

in which each s_i is an expanded instance, and in which $\kappa(V_i)\le\max(\kappa(RX),\kappa(RY))$ for each i.

Proof. Since $R(V\otimes W)=RV\otimes RW$ and $\varrho(V\otimes W)=\varrho V\otimes\varrho W$, and since κ is additive over tensor products, it suffices to consider the case where t is a mere *instance*. If it is an instance of one of the centrals a, a^{-1}, r, r^{-1}, c, the desired (7.3) is given by a naturality diagram such as

$$\begin{array}{ccc}
(V\otimes W)\otimes Z & \xrightarrow{a} & V\otimes(W\otimes Z) \\
{\scriptstyle (\varrho V\otimes\varrho W)\otimes\varrho Z}\downarrow & & \downarrow{\scriptstyle \varrho V\otimes(\varrho W\otimes\varrho Z)} \\
(RV\otimes RW)\otimes RZ & \xrightarrow[a]{} & RV\otimes(RW\otimes RZ).
\end{array}$$

If t is an instance of a map $f: A\to B$ in $\mathscr{A}$, the result is trivial since A and B are reduced. It remains to consider the cases where t is an instance of d or e; we do the first only, the second being entirely dual using the analogues for e of (6.2)–(6.4).

We suppose then that $t: X\to Y$ is $d_Z: I\to Z\otimes Z^*$, and we consider first the case where Z is reduced. If it is $A\in\mathscr{A}$, the matter is trivial, since ϱX and ϱY are each 1. If Z is $V\otimes W$ with V and W reduced, $\varrho I=1$ while $\varrho(Z\otimes Z^*)=1\otimes\varrho(Z^*)=1\otimes u_{VW}$; so a suitable diagram (7.3) is just (6.2) re-written as

$$\begin{array}{ccc}
I & \xrightarrow{\qquad\qquad\qquad d_{V\otimes W}\qquad\qquad\qquad} & (V\otimes W)\otimes(V\otimes W)^* \\
{\scriptstyle 1}\downarrow & & \downarrow{\scriptstyle 1\otimes u_{VW}} \\
I \xrightarrow[d_V]{} V\otimes V^*\cong(V\otimes I)\otimes V^* & \xrightarrow[(1\otimes d_W)\otimes 1]{} & (V\otimes(W\otimes W^*))\otimes V^*\cong(V\otimes W)\otimes(W^*\otimes V^*),
\end{array}$$

where the unnamed isomorphisms are the obvious centrals; clearly no object in the bottom row has greater rank than the last object. The cases $Z=I$ and $Z=A^*$ for $A\in\mathscr{A}$ follow similarly, replacing (6.2) by (6.3) and (6.4).

This leaves only the case where Z is not reduced. But then $\varrho(Z^*)=((\varrho Z)^*)^{-1}$ by the definition of ϱ, and a suitable diagram (7.3) is given by

$$\begin{array}{ccc}
I & \xrightarrow{d_Z} & Z\otimes Z^* \\
{\scriptstyle 1}\downarrow & & \downarrow{\scriptstyle \varrho Z\otimes((\varrho Z)^*)^{-1}} \\
I & \xrightarrow[d_{RZ}]{} & RZ\otimes(RZ)^*,
\end{array}$$

which commutes by the naturality of d. □

Proof of Proposition 7.1. We already have the first assertion. For the second, given a map $X\to Y$ in $F\mathscr{A}$ between reduced objects, write it in the form (7.1) with the t_i expanded instances but with the Z_i not necessarily reduced: we can do this by Proposition 4.1. Then apply Lemma 7.3 with t equal to each t_i in turn, to get a diagram of the form

$$\begin{array}{ccccccccc}
X=Z_0 & \xrightarrow{t_1} & Z_1 & \xrightarrow{t_2} & Z_2 & \longrightarrow\cdots\xrightarrow{t_n} & Z_n=Y \\
{\scriptstyle 1}\downarrow\ \downarrow{\scriptstyle \varrho Z_0} & & \downarrow{\scriptstyle \varrho Z_1} & & \downarrow{\scriptstyle \varrho Z_2} & & \downarrow{\scriptstyle \varrho Z_n}\ \downarrow{\scriptstyle 1} \\
X=RZ_0 & \xrightarrow[t'_1]{} & RZ_1 & \xrightarrow[t'_2]{} & RZ_2 & \longrightarrow\cdots\xrightarrow[t'_n]{} & RZ_n=Y,
\end{array}$$

where each t_i' is the bottom edge of a diagram (7.3). Then the bottom edge of the above diagram is a (longer) composite of expanded instances, in which the new intermediate objects (the V_i of the diagrams (7.3)) all have rank $\leq \max \kappa(RZ_i)$. But this maximum is, by Lemma 7.2, strictly less than $\max \kappa(Z_i)$, unless all the Z_i are already reduced. Hence the result follows by induction. □

8. Proof of the main theorem

We now analyze reduced maps in $F\mathscr{A}$ as tensor products, leading to a direct proof that $F\mathscr{A}$ is isomorphic to $G\mathscr{A}$.

We name some maps in $F\mathscr{A}$. For a map $f: A \to B$ in $\mathscr{A}$, we have f as a map in $F\mathscr{A}$, and we also have $f^*: B^* \to A^*$. We write d_f and e_f for the respective composites

$$I \xrightarrow[d_A]{} A \otimes A^* \xrightarrow[f \otimes 1]{} B \otimes A^*, \qquad B^* \otimes A \xrightarrow[1 \otimes f]{} B^* \otimes B \xrightarrow[e_B]{} I.$$

We denote by ε the composite

$$[\mathscr{A}] \xrightarrow[{[\Phi\mathscr{A}]}]{} [F\mathscr{A}] \xrightarrow[\gamma]{} (F\mathscr{A})(I, I),$$

where this is the γ of Proposition 6.2. We also use the same names for the correspondingly-defined maps in $G\mathscr{A}$. We adopt some conventional meaning for repeated tensor products $X_1 \otimes \cdots \otimes X_n$, such as bracketing from the right if $n > 0$ and I if $n = 0$. Then

Proposition 8.1. *Any map $s: X \to Y$ in $F\mathscr{A}$ between reduced objects is of the form*

$$X \xrightarrow[x]{} V_1 \otimes \cdots \otimes V_m \xrightarrow[h_1 \otimes \cdots \otimes h_m]{} W_1 \otimes \cdots \otimes W_m \xrightarrow[y]{} Y, \tag{8.1}$$

where x and y are central isomorphisms, and where each h_i is either f, f^*, d_f, or e_f for some map f in $\mathscr{A}$, or else $\varepsilon(k)$ for some cycle k in $[\mathscr{A}]$.

Proof. We use induction over the minimum length n of expressions of s as composites (7.1) of expanded instances with the Z_i reduced—such expressions existing by Proposition 7.1. Since the result is trivial if $n = 0$ or 1, it suffices to show that the composite of (8.1) with a single expanded instance $t: Y \to Z$, where Z too is reduced, is again of the form (8.1).

If t is central it can be absorbed in y; so the only cases to consider are those where t is an expanded instance of d_A or of e_A for some $A \in \mathscr{A}$, or else of some map g in $\mathscr{A}$. (We reiterate that d_V and e_V, where V is not a mere object of $\mathscr{A}$, are impossible because Y and Z are reduced.) Note that the central y at the end of (8.1) gives us every flexibility about the order of factors and their bracketing. We leave the reader to fill in the fine detail of simple arguments involving centrals; similar arguments are given in full detail in [10]. Observe that the association of prime factors corresponding to a central (cf. Section 5) is here unique, since no constant primes occur in reduced objects.

If t is an expansion of $d_A : I \to A \otimes A^*$, we can by evident central adjustments move d_A to a position before y in (8.1), and then include it among the h_i as a new h_{m+1}, of the form d_f where $f = 1_A$.

If t is an expansion of some $g : B \to C$ in $\mathscr{A}$, the central y associates the prime factor B of Y with the same prime in some W_i, say W_1. Then h_1 has one of the forms f or d_f for $f : A \to B$ in $\mathscr{A}$; for otherwise W_1 has no prime factors of the form B. Using central adjustments we can move g back before y, and combine it with h_1 to form a new h'_1: when h_1 is f, h'_1 is just gf; and when h_1 is d_f, h'_1 is $(g \otimes 1)d_f$, which is of course d_{gf}; in both cases h'_1 is of one of the desired forms.

There remains the case where t is an expansion of $e_A : A^* \otimes A \to I$. If the prime factors associated by y to A^* and to A lie in the same W_i, let it be W_1. Then h_1 must have the form d_f for some *endomorphism* $f : A \to A$. Carrying e_A back through y, we can combine it with h_1 to make a new h'_1 of the form $e_A c d_f$, which by (6.5) and the naturality of c is $\varepsilon(k)$ where $k = \tau(f) \in [\mathscr{A}]$.

If the prime factors associated by y to A^* and to A lie in different factors W_1 and W_2, there are four possibilities. If h_1 is $f^* : B^* \to A^*$ and h_2 is $g : C \to A$, the carried-back e_A combines with h_1 and h_2 to form $h' = e_A(f^* \otimes g)$, which is e_{fg} by the naturality of e. If h_1 is $f^* : B^* \to A^*$ and h_2 is $d_g : I \to A \otimes C^*$ where $g : C \to A$, we combine e_A with these to get the composite h' given by

$$B^* \cong B^* \otimes I \xrightarrow[f^* \otimes d_g]{} A^* \otimes (A \otimes C^*) \cong (A^* \otimes A) \otimes C^* \xrightarrow[e_A \otimes 1]{} I \otimes C^* \cong C^*,$$

which by (6.1) and the definition of d_g is $g^* f^* = (fg)^*$. The case where h_1 is $d_f : I \to B \otimes A^*$ and h_2 is $g : C \to A$ is similar to this last, transformed by the symmetry c. Finally, if h_1 is $d_f : I \to B \otimes A^*$ and h_2 is $d_g : I \to A \otimes C^*$, we combine e_A with these to get the composite h' given by

$$I \cong I \otimes I \xrightarrow[d_f \otimes d_g]{} (B \otimes A^*) \otimes (A \otimes C^*) \cong B \otimes (A^* \otimes A) \otimes C^* \xrightarrow[1 \otimes e_A \otimes 1]{} B \otimes I \otimes C^* \cong B \otimes C^*,$$

which is d_{fg} by (1.1) and the naturality of d. $\square$

Theorem 8.2. *The function $\Theta : F\mathscr{A} \to G\mathscr{A}$ of* (3.2) *is an isomorphism of compact closed categories.*

Proof. Since Θ by its definition is a strict map of compact closed categories, we have only to show that it is an isomorphism of categories. Since Θ is clearly the identity on objects, it remains to show that it is fully faithful. Since every object of $F\mathscr{A}$ is isomorphic by Proposition 7.1 to a reduced one, it suffices to show that, for reduced X and Y, the map $\Theta_{XY} : (F\mathscr{A})(X, Y) \to (G\mathscr{A})(X, Y)$ is a bijection.

The image under Θ of a central in $F\mathscr{A}$ is the central with the same name in $G\mathscr{A}$. The images under Θ of the above maps $f : A \to B$, $f^* : B^* \to A^*$, $d_f : I \to B \otimes A^*$, $e_f : B^* \otimes A \to I$, where f is a map in $\mathscr{A}$, are the maps with the same names in $G\mathscr{A}$; and it is immediate from the explicit description of $G\mathscr{A}$ in section 3 that each is of the form $(\mathbf{2}, f, 0)$, where $\mathbf{2}$ is the unique involution on $P(A^*\otimes B) = 1^*\otimes 1$ and where $f : \mathbf{2} \to \mathscr{A}$ is the functor corresponding to the map f. Similarly Θ sends $\varepsilon(k) : I \to I$ for $k \in [\mathscr{A}]$ to the $\varepsilon(k)$ in $G\mathscr{A}$, which is at once seen to be $(0,0,k) : I \to I$, where 0 denotes both the empty involution and the unique functor $0 \to \mathscr{A}$, while $k \in [\mathscr{A}]$ is here identified with its image in the free commutative monoid $M[\mathscr{A}]$ on $[\mathscr{A}]$.

To show that Θ_{XY} is surjective it suffices therefore to show that any map $(\theta, p, \lambda) : X \to Y$ in $G\mathscr{A}$ between reduced objects can also be written in the form (8.1). But this is clear: since the involution θ is a coproduct of copies of $\mathbf{2}$, we can express the map $(\theta, p, 0) : X \to Y$ in the form (8.1) by choosing centrals x, y providing the necessary permutations; while $(0,0,\lambda)$ is a tensor product of various $(0,0,k) = \varepsilon(k)$ for $k \in [\mathscr{A}]$; and finally the tensor product of $(\theta, p, 0)$ and $(0,0,\lambda)$ is (θ, p, λ).

It remains to show that Θ_{XY} is injective, by showing that two maps of the form (8.1) in $F\mathscr{A}$ coincide if they have the same image in $G\mathscr{A}$. This comes to showing the following. Let $s : X \to Y$ in $F\mathscr{A}$ be the tensor product, in this order, of maps

$$\begin{gathered} f_1 : A_1 \to B_1, \dots, f_n : A_n \to B_n, \\ g_1^* : D_1^* \to C_1^*, \dots, d(h_1) : I \to F_1 \otimes E_1^*, \dots, e(j_1) : H_1^* \otimes G_1 \to I, \dots, \\ \varepsilon(k_1) : I \to I, \dots, \varepsilon(k_m) : I \to I; \end{gathered}$$

where $d(h)$ and $e(j)$ are now written for d_h and e_j, and where the f_i, g_i, h_i, j_i are maps in $\mathscr{A}$ and the k_i are elements of $[\mathscr{A}]$. Let $s' : X' \to Y'$ be a similar tensor product of primed data, and let $x : X \to X'$, $y : Y \to Y'$ be centrals. Then we are to show that $s'x$ and ys coincide in $F\mathscr{A}$, if they have the same image under Θ.

The commutativity in $G\mathscr{A}$, just at the level of the involutions, shows that the central x must associate the prime factor A_1 of X with a prime factor A_i' in X'; must associate D_1^* with some $D_i'^*$; and must associate H_1^* with $H_i'^*$ and G_1 with G_i' for the *same* i. Exactly corresponding remarks apply to y. So x has the form $x_1 \otimes x_2 \otimes x_3 \otimes x_4 \otimes x_5$, where x_1 is a central from $A_1 \otimes \cdots \otimes A_n$ to $A_1' \otimes \cdots \otimes A_{n'}'$ (note that n' must indeed be n), x_2 one from $B_1^* \otimes \cdots$ to $B_1'^* \otimes \cdots$, x_3 one from $I \otimes \cdots \otimes I$ to $I \otimes \cdots \otimes I$, x_4 one from $(H_1^* \otimes G_1) \otimes \cdots$ to $(H_1'^* \otimes G_1') \otimes \cdots$, and x_5 one from $I \otimes \cdots \otimes I$ to $I \otimes \cdots \otimes I$; and y has a corresponding form $y_1 \otimes y_2 \otimes y_3 \otimes y_4 \otimes y_5$.

By using a central to change the order of the $\otimes$-factors in s', we can suppose that $x_1 = 1$; then commutativity in $G\mathscr{A}$ clearly implies that $y_1 = 1$. Similarly we can suppose $x_2 = 1$, which forces $y_2 = 1$. By another order-change in s' we can suppose $y_3 = 1$; this does not really change x_3, which as a central $I \otimes \cdots \otimes I \to I \otimes \cdots \otimes I$ is already 1, and remains 1 after a permutation of the factors I in the codomain. Similarly we can suppose $x_4 = 1$ and $y_4 = 1$. That $x_5 = 1$ and $y_5 = 1$ is automatic once we know that their domain and codomain have the same number of factors I; that is, that the number m of $\varepsilon(k_i)$-factors in s is the same as the number m' of the $\varepsilon(k_i')$-factors in s'. This follows because commutativity in $G\mathscr{A}$ implies that $k_1 + \cdots + k_m = k_1' + \cdots + k_{m'}'$ in $M[\mathscr{A}]$, whence $m = m'$ because we have the *augmentation* map $M[\mathscr{A}] \to \mathbf{N}$ sending each element of $[\mathscr{A}]$ to 1.

Thus we may suppose $x = 1$ and $y = 1$; and we must show that $s = s'$, given $\Theta s = \Theta s'$

$=(\theta, p, \lambda)$ say. That $f_i = f'_i$, $g_i = g'_i$, $h_i = h'_i$, $j_i = j'_i$ is immediate, since these are recaptured from p. It only remains to show that $\varepsilon(k_1)\otimes\cdots\otimes\varepsilon(k_m)$ in $F\mathscr{A}$ is determined by a knowledge of $k_1+\cdots+k_m \in M[\mathscr{A}]$. But by Proposition 6.1, $\varepsilon(k_1)\otimes\cdots\otimes\varepsilon(k_m)$ is determined by the product $\varepsilon(k_1)\cdots\varepsilon(k_m)$ in the commutative monoid $(F\mathscr{A})(I,I)$; which product is the image of $k_1+\cdots+k_m$ under the monoid-homomorphism $\bar{\varepsilon} : M[\mathscr{A}] \to (F\mathscr{A})(I,I)$ extending $\varepsilon : [\mathscr{A}] \to (F\mathscr{A})(I,I)$. □

9. Related structures

Let us call a compact closed category $\mathscr{B}$ *monoidally strict* if $\mathscr{B}$ is strict as a monoidal category; that is, if $\otimes$ is strictly associative and I a strict unit, with a, l, r identities. Let us call the compact closed category $\mathscr{B}$ *strict* if it is monoidally strict and if, in addition, the isomorphisms u, v, w of Section 6 are identities, so that $(A\otimes B)^* = B^*\otimes A^*$ and so on. Write $F'\mathscr{A}$ [resp. $F''\mathscr{A}$] for the free monoidally strict [resp. strict] compact closed category on $\mathscr{A}$. Then we have canonical functors $F\mathscr{A}\to F'\mathscr{A}\to F''\mathscr{A}$ corresponding to the obvious "maps of theories"; and in this section we show these canonical functors to be equivalences of categories.

First, each of $F'\mathscr{A}$ and $F''\mathscr{A}$ has a description by generators and relations like that of $F\mathscr{A}$ in Section 4. The objects of $F'\mathscr{A}$ are a quotient set of those of $F\mathscr{A}$, by the relations of strict associativity and strict unity; those of $F''\mathscr{A}$ are a still smaller quotient set, and are just the $X_1\otimes\cdots\otimes X_n$ where each X_i is A or A^* for $A\in\mathscr{A}$. In defining the maps of $F'\mathscr{A}$ or $F''\mathscr{A}$, the only difference is that there are no instances of a, l, or r to consider, and that in the case of $F''\mathscr{A}$ we have to add (6.2)–(6.4), with u, v, w identities, as new axioms. In both cases Proposition 4.1 still holds, with a, a^{-1}, r, r^{-1} omitted.

The proofs of Proposition 7.1 and Proposition 8.1 go through unchanged in the case of $F'\mathscr{A}$. In the case of $F''\mathscr{A}$ every object is reduced, and Proposition 7.1 must be replaced by the assertion that every map has the form (7.1) where each t_i is an expanded instance of c, of f, or of d_A or e_A for $A\in\mathscr{A}$; this follows directly from the axioms (6.2)–(6.4) with u, v, w identities. Then in both cases Proposition 8.1 follows as before.

We can define $G'\mathscr{A}$ and $G''\mathscr{A}$ as we defined $G\mathscr{A}$; the objects are those of $F'\mathscr{A}$ and $F''\mathscr{A}$ respectively, but the maps are triples (θ, p, λ) just as in $G\mathscr{A}$. Since $G'\mathscr{A}$ and $G''\mathscr{A}$ are respectively a monoidal strict, and a strict, compact closed category, we get strict functors $\Theta' : F'\mathscr{A}\to G'\mathscr{A}$ and $\Theta'' : F''\mathscr{A}\to G''\mathscr{A}$ analogous to Θ. There are evident strict functors $G\mathscr{A}\to G'\mathscr{A}\to G''\mathscr{A}$, and a commutative diagram

$$\begin{array}{ccccc} F\mathscr{A} & \longrightarrow & F'\mathscr{A} & \longrightarrow & F''\mathscr{A} \\ \downarrow{\scriptstyle\Theta} & & \downarrow{\scriptstyle\Theta'} & & \downarrow{\scriptstyle\Theta''} \\ G\mathscr{A} & \longrightarrow & G'\mathscr{A} & \longrightarrow & G''\mathscr{A}, \end{array}$$

where the functors in the top row are the canonical ones. But Θ' and Θ'' are category-equivalences, like Θ, since Theorem 8.2 for these follows as before from Proposition 8.1. Since the functors in the bottom row are obviously equivalences, so are those in the top row, as required.

10. Remarks on coherence problems

In Section 4 we saw that, for categories with any equational extra structure at all, the explicit determination of the free such $F\mathscr{A}$ on a given category $\mathscr{A}$ involved the solution of a word problem, namely the determination of $\mathfrak{q}\mathscr{A}$ (the commuting diagrams) from $\mathfrak{r}\mathscr{A}$ (the basic ones). In principle there is a prior word problem even in finding the *objects* of $F\mathscr{A}$, if there are axioms expressing equations between objects, as for instance in a strict monoidal category; but in practice the word problem for the objects is usually easy. In any case we can define the *coherence problem* as the explicit determination of $F\mathscr{A}$ in terms of $\mathscr{A}$, and hence the determination of the functor F and the monad on **Cat**.

There is a lack of absoluteness in the above notion of "commuting diagrams": the diagrams in $\mathfrak{q}\mathscr{A}$ depend upon $\mathscr{A}$, and involve its objects and maps. We could get a more "absolute" list of commuting diagrams by specializing to discrete $\mathscr{A}$, or better to an $\mathscr{A}$ which was a coproduct of copies of the unit category **1** and the arrow category **2**.

However MacLane's original coherence theorem [13] for symmetric monoidal categories was formulated quite absolutely, in terms of diagrams of *natural transformations*; not in terms of components of these, and not in terms of the free model $F\mathscr{A}$. The same was true of the Kelly–MacLane coherence result [10] for symmetric monoidal closed categories, the natural transformations there being the generalized ones of [4] between functors of mixed variances.

The first author of the present paper established in [5] and [6] the following connexion, in such cases, between these two formulations. If **1** denotes the unit category, the objects of $F\mathbf{1}$ are just the formal iterates of the structural functors, modulo any equations imposed by the axioms; while the maps of $F\mathbf{1}$ are all the composites of formal expanded instances of the structural natural transformations (*not* components of these), modulo the diagrams $\mathfrak{q}$ which necessarily commute as a consequence of the basic diagrams $\mathfrak{r}$ expressing functoriality, naturality, and the axioms. There is a functor Γ with domain $F\mathbf{1}$, assigning to each formal functor its "arity" and to each formal natural transformation its "type" or "graph". The pair $\mathscr{K}=(F\mathbf{1},\Gamma)$ is a monoid in a certain monoidal category $\mathscr{K}$, whose tensor product represents formal substitution of functors and natural transformations into other functors and natural transformations. There is an action $\circ$ of the monoidal category $\mathscr{K}$ on **Cat**, representing the substitution of actual objects and maps into functors and natural transformations. The free model $F\mathscr{A}$ on $\mathscr{A}$ is then $\mathscr{K}\circ\mathscr{A}$. In such cases, therefore, the determination of F reduces to that of $\mathscr{K}$, and hence of $F\mathbf{1}$; and the coherence problem can be discussed purely in terms of natural transformations, abstracting entirely from any particular $\mathscr{A}$. A monoid $\mathscr{K}$ in **K** was called a *club*, and the monad $UF = \mathscr{K} \circ -$ on **Cat** was said in such cases to arise from a club.

The above results were proved in [5] and [6] only when the natural transformations involved were those of [4], and then only when all composites actually occurring were *compatible* in the sense of [4]. In the case of symmetric monoidal closed categories it had been shown ([10] Theorem 2.2) that no incompatibilities in fact occur; and a very general result of the same nature was proved in [7], thus providing many examples of monads arising from clubs.

There is no need for an equational extra structure on a category to be defined by functors and natural transformations; but many are, and with natural transformations more general than those of [4], in that each variable no longer appears exactly twice. [5] and [6] went on to speculate about possible extensions of the club notion to at least some larger classes of natural transformations, predicting success when these admitted operations of composition and substitution. In fact an extension was later given in [8] to structures defined by purely covariant functors, and by natural transformations which related the variables of the domain and the codomain no longer by a mere pairing-off, but not by a general relation either: instead by a relation that was either a function or the inverse of one.

In another direction there remained the question whether it was possible to extend the club notion to those structures defined by mixed-variance functors and by the "good" natural transformation of [4], in cases where incompatible composites *did* occur. As we saw in Section 1, they do occur for compact closed categories, and this was put forward as a test case. The lack of an *evident* way of composing incompatibles was not a disproof.

Subsequently the first author has given (in the forthcoming [9]) an *absolute* definition of "club", as a monad on **Cat** in which $F\mathscr{A}$ is related in a particular way to $F1$. The explicit determination above of the F for compact closed categories now allows an easy verification (see [9]) that these do *not* arise from a club.

We conclude that those structures whose coherence problem can be discussed purely in terms of suitably-generalized natural transformations are quite special ones, with particularly simple descriptions and particularly good properties; and that in general we must pose the coherence problem in terms of finding $F\mathscr{A}$.

References

[1] M. Barr, ★-Autonomous Categories, Lecture Notes in Math. 752 (1979).

[2] J.H. Conway, On numbers and games (Academic Press, New York-London, 1976).

[3] B.J. Day, Note on compact closed categories, J. Austral. Math. Soc. 24 (Series A) (1977) 309-311.

[4] S. Eilenberg and G.M. Kelly, A generalization of the functorial calculus, J. Algebra 3 (1966) 366-375.

[5] G.M. Kelly, Many-variable functorial calculus I, Lecture Notes in Math. 281 (1972) 66-105.

[6] G.M. Kelly, An abstract approach to coherence, Lecture Notes in Math. 281 (1972) 106-147.
[7] G.M. Kelly, A cut-elimination theorem, Lecture Notes in Math. 281 (1972) 196-213.
[8] G.M. Kelly, *On clubs and doctrines, Lecture Notes in Math.* 420 (1974) 181-256.
[9] G.M. Kelly, An absolute notion of club, (to appear).
[10] G.M. Kelly and S. MacLane, Coherence in closed categories, J. Pure Appl. Algebra 1 (1971) 97-140.
[11] G.M. Kelly and R. Street, Review of the elements of 2-categories, Lecture Notes in Math. 420 (1974) 75-103.
[12] F.W. Lawvere, Metric spaces, generalized logic, and closed categories, Rend. Sem. Matematico è Fisico di Milano 43 (1973) 135-166.
[13] S. MacLane, Natural associativity and commutativity, Rice Univ. Studies 49 (1963) 28-46.