COHERENCE IN NONMONOIDAL CLOSED CATEGORIES

BY

MIGUEL L. LAPLAZA

ABSTRACT. A nonmonoidal closed category is a category with an internal homomorphism functor, left Yoneda natural arrows, unity object and some natural transformations and coherence axioms. The object of this paper is to give a *complete* solution of the coherence problem in this structure: we use a cut-elimination theorem as basic tool to prove that the elementary natural transformations are characterized by their graph (roughly speaking the graph is the type of identification imposed by a natural transformation on the arguments of its domain and codomain).

Introduction. A coherence result can be described either as the study of the conditions of commutativity of a family of diagrams [21], [22, p. 76], or as the description of the diagrams which commute under determined conditions [23], or as the description of a part of the structure of a type of categories which proves the commutativity of a class of diagrams [10, p. 283]. There is also an abstract theory of coherence [8] which can be used as framework for such questions.

In our case the three approaches are equivalent: we have the structure of nonmonoidal closed category with its axioms, a theory of this structure that uses the commutativity of many diagrams and we want to find all the diagrams that *must* commute in that structure. Using an expression of J. R. Isbell [6], "the problem is to what extend elementary diagrams must commute". For technical reasons we study paths rather than diagrams so that our problem reduces to giving a good criterion that establishes whether the composites of two elementary paths are equal.

The categories whose coherence we are going to study were called closed categories by Eilenberg-Kelly [3, p. 421] but we have used in the title of the paper the expression "nonmonoidal closed category" for the sake of clarity because the term closed category has been used as short for different concepts. After this remark we will keep the original terminology of closed category through this paper. Thus following the definition of [3, p. 421] a closed category is a category with internal homomorphism functor, left Yoneda natural arrows, unit object and suitable axioms and natural transfor-

Received by the editors January 23, 1976.

AMS (MOS) subject classifications (1970). Primary 18D15; Secondary 18B15, 18D10, 18D20. Key words and phrases. Closed category, coherence, cut-elimination theorem, graph.

mations, but a closed category is not supposed to be monoidal. On the other hand, and to the effects of the coherence, not all the structure is needed, hence we define the so-called formal closed categories by taking the part of the structure that is relevant for coherence purposes.

The problem of the coherence in closed categories appears explicitly formulated in [23, p. 241] and it can be considered as the "psychological dual" of the coherence in monoidal categories [21]. Moreover the striking resemblance between the axioms (3.5) of [21, p. 33] and CC3 of [3, p. 429] suggested the existence of similar coherence results in both cases.

As we have said we formulate the coherence problem in terms of paths rather than diagrams. If S is a set with the element I we call F(S) the category whose objects and arrows are all the "meaningful formal combinations" of the elements involved in the definition of a formal closed category which contains as objects all the elements of S with I as unit object. If V is a formal closed category, any map from S into Ob V that preserves the unit object can be lifted to a functor from F(S) into V: two arrows of S are said to be equivalent when their images by any such functor are equal. The basic aim of the paper is to prove that two arrows of F(S) with the same domain and codomain are equivalent if and only if they have the same graph in the sense of [4]. We must recall that we are dealing with natural arrows in a generalized sense where arguments of the domain and codomain can be identified and the graph can be defined intuitively as the pattern of this identification. The paper studies the category $Shap_v = Shap$, of the shapes with respect to a formal closed category V, defined as the category F(S)when S = Ob V: the identity map of S originates a strict closed functor $| \cdot |$: **Shap** \rightarrow V and the coherence theorem of the paper proves that for f, g: $A \rightarrow B$ arrows of Shap the equality of the graphs implies |f| = |g| and the solution of the coherence problem is a straightforward consequence of this fact.

We are going to mention two other approaches to the solution of this problem that as long as we know have not been carried out yet. There exists an embedding of a closed category into a nonassociative monoidal closed category [1], [2], [9] that reduces the problem to that of the coherence in a nonassociative monoidal closed category; this is an open question although for the nonassociative monoidal case the solution is known [16]. Another possibility is to use the embedding of a closed category into a biclosed category [19] and the problem is reduced to the coherence in a biclosed category problem which has not been resolved completely [23, p. 241], [14, p. 116], [8, p. 144].

The paper is divided into four sections. In the first section we detail the definitions and the basic concepts. In the second section we study up to equivalences the structure of the arrows of the category of shapes of a formal

closed category and we prove the cut-elimination theorem. The third section contains some preliminary results and the coherence theorem. In the last section we give the general solution and some remarks on the results.

We assume some familiarity with the concepts of graph as used in [4], [8], [9] or [11], natural transformation in the generalized sense of [4] or [24, pp. 214–218] and closed category [3]. Throughout the paper we will adopt the following conventions:

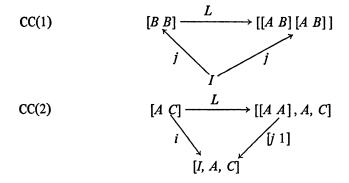
- (a) We will omit as many subindices as can be recovered from the context.
- (b) We will use the same symbol to denote a natural transformation and its inverse if it exists in the ordinary sense or in the formal sense of §1; the context will make clear the exact meaning of the symbol.
- (c) We will use the symbol [] to denote the internal homomorphism functor and sometimes we will use a comma to make clear its two components.
- (d) The symbol [A, B, ...] is defined recursively by [A, B, ...] = [A[B...]]; $[A_1, A_2, ..., A_n, B] = B$ if n = 0.
- (e) The identity arrows will be denoted by the symbol 1 with a subindex if necessary.
 - (f) We will follow [24] for all undefined concepts.

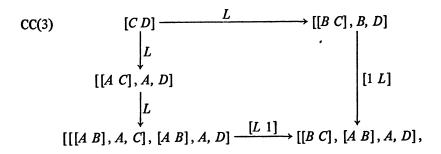
The author is deeply indebted to Professor S. Mac Lane for his advice during the elaboration of this paper.

- 1. **Definitions and basic concepts.** A formal closed category is a category V with the following structure:
 - (1) A functor []: $V^{op} \times V \rightarrow V$.
 - (ii) An object I of V.
 - (iii) A natural isomorphism, $i_A: A \rightarrow [I A]$.
 - (iv) A natural transformation, $j_A: I \to [A \ A]$.
 - (v) A natural transformation,

$$L_{A,B,C}$$
: $[BC] \rightarrow [[AB][AC]]$.

These data satisfy the axioms expressed by the commutativity of the following types of diagrams:





CC(4)
$$[B \ C] \xrightarrow{L} [[I \ B], I, C]$$

$$[B, I, C]$$

$$\mathrm{CC}(5) \qquad \qquad i_I = j_I \colon I \to [I \ I] \, .$$

A closed category in the sense of Eilenberg-Kelly [3, p. 428] is essentially [9, p. 51] a formal closed category with the additional axiom that the map $f \rightsquigarrow [1, f]j$ be an isomorphism from V(A, B) onto V(I, [A B]).

Let \mathcal{C} be the category whose objects are the small categories A with the following structure:

- (i) An object I.
- (ii) A map []: Ob $A \times Ob A \rightarrow Ob A$ and families of arrows (with no assumptions of naturality),

$$i_{A} \colon A \to [I A].$$

$$i_{A} \colon [I A] \to A,$$

$$j_{A} \colon I \to [A A],$$

$$L_{A,B,C} \colon [B C] \to [[A, B], A, C],$$

where A, B and C are objects of A.

- (iii) A partial map [] from Arr $A \times Arr A$ onto Arr A that satisfies the conditions given below.
 - (iv) If $1: A \rightarrow A$ is the identity, [1, 1] is defined and is the identity.
- (v) If $f: X \to Y$ is either one element of the family referred to in (ii), or an identity, or an element of the image of [] and 1: $A \to A$ is an identity, then

$$[f, 1]: [YA] \rightarrow [XA], [1, f]: [AX] \rightarrow [AY],$$

are defined.

The arrows of \mathcal{C} are the functors which preserve strictly the above structure.

Note that if f and g are arrows of a category A of \mathcal{C} the symbol [f, g] is not always defined. We will adopt the convention of representing by [f, g] any of the arrows [f, 1][1, g] and [1, g][f, 1]: whenever this convention is used the results will not depend upon the choice or will be true for any choice, and this will be also the case when [f, g] is defined in A.

Let S_I be the category of sets with fixed point I and let $G: \mathcal{C} \to S_I$ be the forgetful functor which takes each category onto its objects: this functor has a left adjoint, F, as can be proved by means of the Freyd's adjoint functor theorem. We are going to give a construction of F: an analogous construction was used in [18, p. 41] or [17] and the same idea was present in [16, p. 108] although this last situation was simple enough to allow the omission of the description of the arrows.

Let S be an object of S_I . The objects of F(S) are the elements of the free []-algebra over S: we will assume that $S \subset Ob F(S)$. Let now G be the graph (in the usual sense of graph theory) whose edges are all the formal expressions of type:

$$\begin{split} i_{A} \colon A \to & \big[I A \big], \\ i_{A} \colon \big[I A \big] \to A, \\ j_{A} \colon I \to & \big[A A \big], \\ L_{A,B,C} \colon \big[B C \big] \to & \big[\big[A B \big], A, C \big], \\ 1 \colon A \to A, \end{split}$$

where A, B and C are elements of F(S). Define $\mathcal K$ as the free []-algebra over the edges of $\mathcal G$ and d_0 , d_1 : $\mathcal K \to \operatorname{Ob} F(S)$ as the two maps extending the graph structure of $\mathcal G$ and such that

$$f: A \to B, g: C \to D \Rightarrow [f, g]: [BC] \to [AD].$$

Finally take as \mathcal{F} the set of all the edges of \mathcal{K} in whose "[]-composition" there is exactly one factor of type i, j, or L and all the others, if any, are identities: the elements of \mathcal{F} will be called instances of i, j or L, respectively. The arrows of F(S) are the paths with edges in \mathcal{F} and the concatenation for composition [24, p. 50]; the identities are the paths with one vertex and no edges. It is immediate to complete the structure of F(S) to be an object of \mathcal{C} . Throughout any "unlabeled" adjunction will be supposed to be the adjunction between F and G as above. We will define the *length* of an arrow of F(S) as the number of edges in its path.

Let $m: S \to \text{Ob } A$ be an *I*-preserving map and $m': F(S) \to A$ the arrow defined by the action of the adjunction on m. The action of m' on the objects and arrows of F(S) can be described intuitively as the result of the action of

m on the occurrences of objects of S in the objects and in the "labels" of the edges of the paths defining the arrows of F(S).

We define that two arrows, $f, g: X \to Y$, of F(S) are equivalent if for any formal closed category V and the adjoint $m': F(S) \to V$ of any I-preserving map $m: S \to Ob V$ we have m'(f) = m'(g). In more intuitive words we can reformulate the concept of equivalence as follows: f and g are formal composites of symbols that make sense as arrows in any formal closed category if we give values into the objects of that category to the occurrences of elements of S in the label of the expressions of the arrows f and g; f and g are equivalent if they take the same value in any such process. Note that the axioms of the formal closed categories imply the equivalence of some arrows in any category F(S). Thus the axiom CC(1) implies that two arrows of type $j_{[A \ B]}$ and $L_{A,B,B}j_B$ are equivalent in any category F(S). By the same reason the naturality of i, j, L and the functoriality of [I] produce equivalences in F(S).

The aim of this paper is to find "good" conditions for the equivalence of arrows in the categories of type F(S). To do that we have to define the category **Graph** which is very closely related to the category G of [11, p. 103].

Let op: $\{+, -\} \rightarrow \{+, -\}$ be the map defined by op(+) = -, op(-) = +. The action of op can be extended to the finite sequences of elements of $\{+, -\}$ by the definition:

$$\operatorname{op}(\emptyset) = \emptyset,$$

 $\operatorname{op}(a_1 \dots a_t) = (\operatorname{op}(a_1) \dots \operatorname{op}(a_t)).$

If A and B are finite sequences of elements of $\{+, -\}$ we will define [A B] as the concatenation of op(A) and B. The finite sequence $[A \ B]$ can be considered as a finite sequence of "points" in A or B each one of them with variance + or -; by using this terminology we can define a graph, $g: A \rightarrow B$, from A onto B as a fixed-point-free involution of [A B] such that it maps any point onto a point of opposite variance. The best way of defining the composition of two graphs, $f: A \to B$, $g: B \to C$, is to consider f and g as a set of links between points; then the composite gf: $A \rightarrow C$ is the graph of all the links between points x and y of $[A \ C]$ for which there is a finite sequence, $x = x_1, x_2 \dots x_t = y$, such that x_i and x_{i+1} are linked alternatively by f and g. The computation of the products of graphs can be done very easily by representing them as "geometrical graphs" [7, pp. 98-101], [8, p. 130]. It is routine to check that we have defined a category that we will represent by **Graph.** Moreover **Graph** is a closed category: we take $I = \emptyset$; the internal homomorphism functor [] has been defined already for the objects, and for arrows f and g the graph [f, g] is just the union of graphs if they are considered as sets of links among points; the isomorphism i is the identity and the natural transformations j and L consist of the links between identical points of the repeated arguments. It is very easy to check that **Graph** is a closed category hence a formal closed one.

If S is a set containing I there is an obvious map $g: S \to \text{Ob Graph}$ given by g(x) = (+) for $x \neq I$ and $g(I) = \emptyset$. The adjunction takes g onto a functor denoted by graph_S or simply by graph: $F(S) \to \text{Graph}$. On the objects graph substitutes the objects by the ordered formal variance of the arguments: we say formal in the sense that for it to be the variance in the usual sense F(S) would have to be formal closed so that [] would be a functor. For $f: X \to Y$ an arrow of F(S), graph(f) is the formal analogue of the concept of graph of [4, p. 372]. Suppose that V is a formal closed category, $m: S \to \text{Ob V}$ an I-preserving map and $m': F(S) \to \text{V}$ the transform of m by the adjunction. It is easy to check that for any arrow $f: X \to Y$ of F(S), graph(f) gives the graph of $m'(f): m'(X) \to m'(Y)$ in the sense of [4, p. 372].

If V is a formal closed category, the category of shapes of V, denoted by $Shap_V$ or simply by Shap, is the category $F(Ob\ V)$: the objects of Shap will be called shapes. The identity map of $Ob\ V$ defines through the adjunction a functor, $|\ |_V$: $Shap \to V$ which in fact is the evaluation at V of the counit of the adjunction. Thus we have two functors:

$$\big| \big|_V \colon Shap \to V,$$
 graph_V: Shap $\to Graph$.

The core of the paper is the proof of the implication,

$$graph(f) = graph(g) \Rightarrow |f| = |g|,$$

for arrows f and g of **Shap** with the same domain and codomain; the solution of the coherence problem is an immediate consequence as we will see in the last part of the paper. From an intuitive viewpoint the above implication says that the "real" value of any formal combination of elements of the structure of a formal closed category depends only upon the formal expression of the domain and codomain and the graph of the formal combination.

The following definitions will be used throughout the paper and refer to a formal closed category V and its category of shapes.

A central arrow of **Shap** is an identity or any arrow whose edges are instances of i.

A constant shape is any object of **Shap** which involves only I in its expression.

A constant arrow of **Shap** is any arrow which has only constant shapes in its vertices.

Let N be the set of nonnegative integers. Besides the length of arrows of F(S) the following parameters will be used for induction purposes:

The weight is a map w: Ob Shap $\rightarrow \{0, 1\}$ which is defined by the axioms:

- (i) w(I) = 0,
- (ii) if $X \in \text{Ob } V \{I\}, w(X) = 1$,
- (iii) $w[A \ B] = |w(A) w(B)|$.

The basic property of the weight is that the existence of an arrow $f: X \to Y$ in **Shap** implies w(X) = w(Y) as can be proved immediately by induction on length(f).

The rank is a pair of N-valued maps defined over the objects and arrows of Shap. The map rank: Ob Shap \rightarrow N is defined recursively by the axioms:

- (i) rank(I) = 0,
- (ii) if $X \in \text{Ob } V \{I\}$, rank(X) = 1,
- (iii) if A is a constant shape, rank[A X] = rank X,
- (iv) if A is not a constant shape,

$$rank[A X] = rank A + rank X + 1.$$

For the arrows of Shap the rank is defined by

$$rank(f: A \rightarrow B) = rank A + rank B$$
.

Note that the equivalence of two arrows implies the equality of their ranks because the rank only depends upon the domain and codomain of the arrow. This rank is analogous but different from that of [21, p. 33], and [11, pp. 120–121].

2. The structure of the arrows of Shap and the cut-elimination theorem. Throughout this section we will suppose that V is a formal closed category and Shap the category of the shapes of V.

By induction on the length of the arrows of Shap and using straightforward arguments it is easy to prove the following two propositions:

PROPOSITION 1. Any arrow of Shap of type $f: I \rightarrow [A \ B]$ is equivalent to an arrow of type

$$[1g]j: I \rightarrow [AA] \rightarrow [AB],$$

where length g < length f.

PROPOSITION 2. Any arrow of Shap of type $f: [A \ B] \rightarrow I$ is equivalent to an arrow of type

$$i[g,h]: [AB] \rightarrow [II] \rightarrow I.$$

The next proposition requires a not so direct proof.

PROPOSITION 3. Any arrow of Shap of type $f: I \to I$ is constant, that is, any vertex of the path defining f is constant.

PROOF. Suppose that there is an arrow of type $f: I \to I$ which is not constant. This means that there exist arrows in **Shap** such that $f: I \to X \to I$, where X is not constant. We can take f to be such an arrow so that the

number of occurrences of objects of V in the expression of X is minimal: recall that Ob Shap is the free []-algebra over Ob V and that Ob $V \subset V$ Ob Shap. Taking into account the basic property of weight we have:

$$w(X) = w(I) = 0,$$

so that $X \not\in \text{Ob V}$, hence $X = [A \ B]$ for some A and B; moreover we can suppose that A is not constant because otherwise there exist arrows in Shap of type

$$[A B] \rightarrow B, \qquad B \rightarrow [A B],$$

and that proves the existence in Shap of an arrow of type

$$I \to X = [A \ B] \to B \to [A \ B] \to I$$

and this fact goes against the minimality of X. By Proposition 1 the arrow f is equivalent to an arrow of type

$$I \rightarrow [A A] \rightarrow I$$
,

and by Proposition 2 there are arrows in Shap of type $I \to A$, $A \to I$ which by composition give an arrow of type

$$I \rightarrow A \rightarrow I$$
.

where A is not constant and this is a contradiction to the minimality of X = [A B].

Let A be an object of **Shap**: we can use instances of i to obtain shapes with fewer occurrences of I: the result of doing this operation as many times as possible together with the compositive of all the instances is defined to be a reduction of A. A more formal definition is the following: a *reduction* of a shape X is any one of the arrows $\operatorname{red}_X \colon X \to \operatorname{red} X$ and $\operatorname{red}_X \colon \operatorname{red} X \to X$ obtained by the application of the following rules:

- (i) If $A \in Ob V$, 1: $A \rightarrow A$ is a reduction of A.
- (ii) If $a: A \to \text{red } A$, $a: \text{red } A \to A$ are reductions of A, then

$$a i: [IA] \rightarrow A \rightarrow red A$$

and

$$i \ a : \operatorname{red} A \to A \to [I \ A]$$

are reductions of [I A].

(iii) If 1: $A \rightarrow A$ and b: $B \rightarrow \text{red } B$, b: red $B \rightarrow B$ are reductions of A and B respectively and $A \neq I$, then

$$[1, b]: [A B] \rightarrow [A, red B]$$

and

$$[1, b]: [A, red B] \rightarrow [A B]$$

are reductions of [A B].

(iv) If $a: A \to \text{red } A$, $a: \text{red } A \to A$ and $b: [\text{red } A, B] \to \text{red}[\text{red } A, B]$, $b: \text{red}[\text{red } A, B] \to [\text{red } A, B]$ are reductions of A and [red A, B] respectively and a is not the identity, then

$$b[a, 1]: [A B] \rightarrow [red A, B] \rightarrow red[red A, B]$$

and

$$[a, 1]b : red[red A, B] \rightarrow [red A, B] \rightarrow [A B]$$

are reductions of [A B].

We will say that a shape A is reduced if 1: $A \rightarrow A$ is a reduction of A or, equivalently, if no occurrence of I in A can be eliminated by instances of i.

It is immediate that the reduction of any shape exists and it is unique. Moreover if $A \rightarrow \text{red } A$ is a reduction of A the shape red A is a reduced shape.

The concept of reduction is the basic tool in the proof of the following proposition that studies the equivalence of central arrows. Similar concepts denoted reduction and normalization were used in [18, pp. 46 and 59].

PROPOSITION 4. If $f, g: A \to B$ are central arrows in Shap and $A \to \text{red } A$, $B \to \text{red } B$ are reductions of A and B respectively, then f and g are equivalent and red A = red B.

PROOF. Suppose that $h: A \to B$ is a central arrow of **Shap** and $a: A \to \operatorname{red} A$, $b: B \to \operatorname{red} B$ are reductions of A and B respectively. It is easy to prove by induction on length of h that $\operatorname{red} A = \operatorname{red} B$ and that the arrows

$$bh: A \to B \to red B$$
, $a: A \to red A$.

are equivalent. Taking into account that the "value" of a reduction in a formal closed category is an isomorphism the above fact implies that h is determined up to equivalences by A and B; this proves the proposition.

If X_1, \ldots, X_n , A and B are objects of **V** and $n \ge 1$ the symbol $L_{X_1, \ldots, X_n, A, B}$ is defined by the axioms:

(i) $L_{X_n,A,B}$: $[A \ B] \rightarrow [[X_n \ A], X_n, B]$ is the ordinary left Yoneda arrow.

(ii) If
$$L_{X_1, \dots, X_n, A, B} : [A, B] \rightarrow [U, V]$$
, then

$$L_{X_1,X_2\cdots X_n,A,B} = L_{X_1,U,V}L_{X_2\cdots X_n,A,B}\colon \left[A\ B\right] \to \left[U\ V\right] \to \left[\left[X_1\ U\right],X_1,V\right].$$

Intuitively we can define,

$$L_{X_1\ldots X_n,A,B}=L_{X_1}L_{X_2}\ldots L_{X_n,A,B},$$

where each L_{X_i} is a left Yoneda arrow and the subindices have to be completed so that the composite exists. Moreover an analogous definition can be given for **Shap** and in fact we will use this concept later.

By induction on n and straightforward diagram-chase arguments (whose only difficulty is the size of the drawing) we can prove the commutativity of the following type of diagrams:

$$[A B] \xrightarrow{L} [[C A], C, B]$$

$$\downarrow L \qquad \qquad \downarrow [1 L]$$

$$[[Y_1 \cdots Y_r A], Y_1 \cdots Y_r B] \qquad [[C A], [Y_1 \cdots Y_r C], Y_1 \cdots Y_r B]$$

$$\downarrow L \qquad \qquad \downarrow [L 1]$$

$$[[[Y_1 \cdots Y_r C], Y_1 \cdots Y_r A], [Y_1 \cdots Y_r C], Y_1 \cdots Y_r B]$$

This fact is useful if one wants to write down in complete detail the proofs of some of the next propositions: nevertheless it will not be used explicitly in this paper because we have omitted some computations.

PROPOSITION 5 (CUT-ELIMINATION THEOREM). Any arrow $u: X \to Y$ of Shap is of one of the following types:

Type I. An arrow equivalent to a central arrow.

Type II. An arrow equivalent to an arrow of the form

$$c'[f,g]c: X \rightarrow [A B] \rightarrow [A' B'] \rightarrow Y,$$

where c and c' are central arrows and f is not constant.

Type III. An arrow equivalent to an arrow of the form

$$c'[h\ 1]L[1\ g]c\colon X\to [A\ B]\to [A\ C]$$
$$\to [[X_1\ldots X_r,A],X_1\ldots X_r,C]\to [D,X_1\ldots X_r,C]\to Y,$$

where c is central, X, D and at least one of the X_i 's are not constant and c' is the identity or a central arrow of the form

$$[D, X_1 \ldots X_r, C] \rightarrow [K_1 \ldots K_r, D, X_1 \ldots X_r, C]$$

where the K_i 's are constant.

Type IV. An arrow equivalent to an arrow of the form

$$c'[1g]jc: X \to I \to [AA] \to [AB] \to Y,$$

where c and c' are central and A is not constant.

PROOF. We will omit a detailed exposition of the proof that requires more patience than ingenuity and we will limit ourselves to giving a general outline: one has to use induction on the length of u and rather straightforward diagram-chase arguments, being patient enough to detail all the cases and subcases which appear. Although the best way of organizing this induction can depend on the personal characteristics of the reader we will outline the method that we have followed.

The case length(u) = 1 is immediate. In the general case we studied separately the case of X or Y being a constant shape because then we can use

Propositions 1, 2 or 3. In the remaining cases we decompose u as

$$u = hf: L X \rightarrow Y \rightarrow Z$$

where length(h) = 1, so that f satisfies the induction hypothesis and we can assume that f is of one of the types given in the proposition. After this we subdivided the case into the consideration of h being central or noncentral and we considered the possibilities of the types of f and form of h.

A crucial consequence of the proposition is that the rank of the arrows f and g which appear in Type II, the rank of h and g which appear in Type III and the rank of g which appears in Type IV is lower than the rank of g this fact will be used for induction purposes.

An alternative formulation of the above proposition would be the following: define a constructible arrow of Shap as an arrow satisfying the proposition and then our cut-elimination theorem is equivalent to the proposition that the class of constructible arrows is closed for composition. Under this formulation the above proposition appears more similar to the so-called cut-elimination theorems that appear elsewhere.

The cut-elimination technique of substituting the ordinary composition of arrows for another type of composition more suitable for induction purposes seems to be used for the first time for coherence problems in [11] and according to [11, p. 101] and to [8, p. 142] it was borrowed from Gentzen [5] via Lambek [13], [14]. As we have pointed out after Proposition 5, our cut-elimination theorem can be put in a form closely related to some propositions [11, p. 119], [12, p. 23], [20, p. 185] which appear in the framework of the theory of symmetric closed monoidal categories: there is a deep reason for this fact because as we will see in the last part of the paper the category Shap can be used as a kind of "skeleton" of a closed category and any closed category can be embedded structure-preservingly in a biclosed category [19].

PROPOSITION 6. If $[P \ Q]$ is a reduced shape, any central arrow of type c: $[P \ Q] \rightarrow R$ is equivalent to an arrow of the form

$$[c_1 \dots c_t]i \dots i: [PQ] \rightarrow [I, P, Q] \rightarrow \dots$$

$$\rightarrow [I, I \dots I, P, Q] \rightarrow [K_1 \dots K_{t-2}, P', Q'],$$

where the ci's are central arrows and the Ki's are constant shapes.

PROOF. Induction on length(c). The case length(c) = 1 is trivial taking into account that $[P \ Q]$ is reduced; in the general case and by using Proposition 4 it is sufficient to check that the composition of an arrow which satisfies the proposition with an instance of i gives an arrow equivalent to an arrow of the same type and this is immediate.

An analogous proposition holds for central arrows of type $c: R \to [P \ Q]$ where $[P \ Q]$ is reduced: it is enough to take the "formal inverse" of c.

3. The coherence theorem. Throughout this section V will be a formal closed category, Shap the category of shapes of V and graph: Shap \rightarrow Graph and $| \cdot |$: Shap \rightarrow V the functors defined in §1.

Recall that the objects of **Shap** are the elements of the free []-algebra over Ob V so that they can be thought to be ordered sequences of objects of V combined with symbols [and]: we will call points of a shape the occurrences of the objects of V in the expression of the shape. The functor graph: **Shap** \rightarrow **Graph** takes the shapes onto ordered sequences of elements of $\{+,-\}$ in such a way that to each point of X that is not an occurrence of I corresponds a point of graph(X). If $\hat{f}: X \rightarrow Y$ is an arrow of **Shap**, graph(f) is a set of links between points of graph(X) and graph(Y) and via the correspondence just mentioned we can also suppose that graph(f) is a set of links between the points of X and Y. Then the following definition makes sense: if $f: X \rightarrow Y$ is an arrow of **Shap** we will say that a set of points of X and Y is closed for graph(f) if any point of the set is an occurrence of I or is linked by graph(f) with another point of the same set.

PROPOSITION 7. Let $f: [A_1, \ldots, A_r] \rightarrow [B_1, \ldots, B_s]$ be an arrow of **Shap**. Then,

(i) if A_i is closed for graph(f), f is equivalent to an arrow of type

$$h[1 \ldots g \ldots 1]: [A_1 \ldots A_i \ldots A_r] \rightarrow [A_1 \ldots I \ldots A_r] \rightarrow [B_1 \ldots B_s],$$

(ii) if B_i is closed for graph(f), f is equivalent to an arrow of type

$$[1 \ldots p \ldots 1]q: [A_1 \ldots A_r] \rightarrow [B_1 \ldots I \ldots B_s] \rightarrow [B_1 \ldots B_j \ldots B_s].$$

PROOF. If the A_i or B_j referred to in the hypothesis of the proposition are constant shapes the proposition is immediate because there are central arrows of type

$$\begin{aligned} A_i &\to I \to A_i, \quad I \to A_i \to I, \\ B_i &\to I \to B_i, \quad I \to B_i \to I, \end{aligned}$$

that are equivalent to the identity. Thus it is easy to reduce the proposition to the case of the domain and codomain of f being reduced shapes and to prove it by induction on rank(f). The case rank(f) = 0 is trivial and in the general case the cut-elimination theorem can be used to prove the proposition by means of a straightforward argument.

We will define a *flat* arrow as an arrow $f: X \to Y$ of **Shap** which satisfies the following axioms:

(i) The shapes X and Y are reduced.

(ii) If $X = [A_1, \ldots, A_r]$, r > 1 and $A_i \neq I$, then the set of points of A_i is (iii) If $Y = [B_1, \ldots, B_s]$, s > 1 and $B_j \neq I$, then the set of points of B_j is not closed for graph(f).

Intuitively speaking a flat arrow is an arrow with reduced domain and codomain which cannot be decomposed properly by means of Proposition 7.

The next proposition is a simplified version of the cut-elimination theorem valid only for flat arrows.

PROPOSITION 8. If $u: X \to Y$ is a flat arrow of Shap then u is of one of the following types:

Reduced type I. An arrow equivalent to the identity.

Reduced type II. An arrow equivalent to an arrow of the form

$$[g, h]: [A B] \rightarrow [C D].$$

Reduced type III. An arrow equivalent to an arrow of the form

$$[q \ 1]L[1 p]c: X \to [A B] \to [A C] \to [[X_1 \dots X_r, A], X_1 \dots X_r, C]$$

 $\to [D, X_1 \dots X_r, C],$

where $r \ge 1$ and c is the identity or $c = i: X \rightarrow [I X]$.

Reduced type IV. An arrow equivalent to an arrow of the form

$$[1 \ q]j: I \rightarrow [A \ A] \rightarrow [A \ B].$$

PROOF. By Proposition 5, u can be of four different types and it will be sufficient to prove the proposition for everyone of them.

Type I. The arrow u is central and by Proposition 4, X = Y and f is equivalent to the identity hence f is of reduced type I.

Type II. The arrow u is equivalent to an arrow of the form

$$c'[f,g]c:X\rightarrow [A\ B]\rightarrow [A'\ B']\rightarrow Y,$$

where c and c' are central and f is not constant. Proposition 3 implies that A and A' can not be both constant and because u is flat A is constant if and only if A' is constant: then Proposition 6 proves almost immediately that u is of reduced type II.

Type III. The fact of y being reduced implies that if u is of Type III it is equivalent to an arrow of the form

$$[h \ 1]L[1 \ g]c: X \to [A \ B] \to [A \ C]$$
$$\to [[X_1 \dots X_r, A], X_1 \dots X_r, C] \to [D, X_1 \dots X_r, C] = Y,$$

where r > 1 and c is central. By Propositions 4 and 6, the functoriality of [] and the naturality of L and i in formal closed categories we can assume that c is the identity or c = i: $X \rightarrow [I X]$, so that u is of reduced type III.

Type IV. If u is of Type IV the fact of X being reduced implies that u is equivalent to an arrow of the form

$$c'[1 \ q]j: I \rightarrow [A \ A] \rightarrow [A \ B] \rightarrow Y,$$

where c' is central and A is not constant. Proposition 6 allows us to prove easily that we can suppose that c' is the identity so that u is of reduced Type IV.

PROPOSITION 9. Let $f: X \to [U_1 \dots U_s, V_1 \dots V_t, V]$ be an arrow of Shap with $s \ge 0$, t > 1. Suppose that any point in one of the shapes V_i is linked by graph(f) with another point in one of the shapes V_j and suppose moreover that the set of points of any V_i is not closed for graph(f). Then f is equivalent to an arrow of type

$$[1\ldots 1,g]b:X\to [U_1\ldots U_s,V]\to [U_1\ldots U_s,V_1\ldots V_s,V],$$

where g is of type

$$g = [h, 1 \dots 1]L i: V \rightarrow [I V]$$

$$\rightarrow [[V_2 \dots V_t, I], V_2 \dots V_t, V] \rightarrow [V_1 \dots V_t, V].$$

PROOF. We will give only the general outline of the proof that is a long induction on t + rank(f). The case t + rank(f) = 2 is trivial and for the general case it is convenient to consider separately the three following subcases:

- (a) There is a q such that $1 \le q < t$ and any point in $\{V_1, \ldots, V_q\}$ is linked by graph(f) with another point in $\{V_1, \ldots, V_q\}$: it is very easy to use twice the induction hypothesis to prove the proposition.
- (b) The arrow f is flat: one has to consider separately the reduced types given by Proposition 8. When f is of reduced type I, II or IV the proof of the Proposition is straightforward. When f is of reduced type III it is equivalent to an arrow of form

$$X \to [A \ B] \to [A \ C] \to [[X_1 \dots X_r, A], X_1 \dots X_r, C]$$

$$\to [D, X_1 \dots X_r, C] = [U_1 \dots U_s, V_1 \dots V_t, V],$$

where $X \to [A \ B]$ is the identity or $i: X \to [I \ X]$, and the proof of the proposition can be done easily by studying separately the following situations:

- (b)(i) s = 0, (b)(ii) $s \neq 0$, C = V or $C = [V_1 \dots V_t, V]$ for i > 1, (b)(iii) $s \neq 0$, $C = [V_1 \dots V_t, V]$, (b)(iv) $s \neq 0$, $C = [U_i \dots U_s, V_1 \dots V_t, V]$ for i > 1, (b)(v) $V = [X_p \dots X_r, C]$, for $p \geq 1$.
- (c) In the general case it is easy to prove that one can suppose without loss of generality that X is reduced. Moreover it is easy to prove that if an arrow

of f satisfies the proposition so does cf when c is an instance of $i: B \to [I B]$, and that allows us to reduce the proof of the proposition to the case of $[U_1 \ldots U_s, V_1 \ldots V_t, V]$ being a reduced shape and either we are in case (b) or we can use Proposition 7 to apply the induction hypothesis to prove the proposition.

PROPOSITION 10 (COHERENCE THEOREM). If $f, g: X \to Y$ are arrows in Shap such that graph(f) = graph(g), then |f| = |g|.

PROOF. We will use induction on rank(f). If we take into account that the image by $| \ |$ of any reduction is an isomorphism we can assume without loss of generality that X and Y are reduced.

The case $\operatorname{rank}(f) = 0$ is trivial. For the general case we have to remark that the relation $\operatorname{graph}(f) = \operatorname{graph}(g)$ implies that f is flat if and only if g is flat. Consequently if f is not flat we can use Proposition 7 to find similar decompositions of f and g in which we can use immediately the induction hypothesis. Thus we can suppose that f and g are flat.

A case to be considered separately is when

$$Y = [U_1 \ldots U_s, V_1 \ldots V_t, V]$$

and any point of a V_i is linked by graph(f) = graph(g) with another point in a V_j : as f and g are flat no one of the V_i is closed for graph(f) = graph(g) and we can use Proposition 9 to find decompositions of f and g where we can use easily the induction hypothesis.

In the remaining cases we can examine the reduced types given in Proposition 8 and the proof turns out to be straightforward when f or g is of reduced type I, II or IV. Thus the only case to be studied is when f and g are of reduced type III and within trivial permutation of the data we have to consider the two following possibilities:

$$f = [f'', 1 \dots 1]Lf': [A B] \rightarrow [A C]$$

$$\rightarrow [[X_1 \dots X_r, A], X_1 \dots X_r, C] \rightarrow [D, X_1 \dots X_r, C],$$

$$g = [g'', 1 \dots 1]Lg': [A B] \rightarrow [A C]$$

$$\rightarrow [[X_1 \dots X_r, A], X_1 \dots X_r, C] \rightarrow [D, X_1 \dots X_r, C],$$

$$f = [f'', 1 \dots 1]Lf': [A B] \rightarrow [A C]$$

$$\rightarrow [[X_1 \dots X_r, A], X_1 \dots X_r, C] \rightarrow [D, X_1 \dots X_r, C],$$

$$g = [g'', 1 \dots 1]Lg': [A B] \rightarrow [A, X_{s+1} \dots X_r, C]$$

$$\rightarrow [[X_1 \dots X_s, A], X_1 \dots X_s, X_{s+1} \dots X_r, C]$$

$$\rightarrow [D, X_1 \dots X_r, C].$$

In the case (i) we can use immediately the induction hypothesis. In the case (ii) it is easy to deduce from the structure of f and g and the relation graph(f) = graph(g) that the points of $\{X_{s+1} \dots X_r\}$ are closed for graph(f) and this case was studied already.

As we will see in the next section this proposition is the key for an almost immediate solution of the coherence problem and this is the reason for the name of "coherence theorem".

4. The solution of the coherence problem. Throughout this section S will be a set including I as an element and F(S), graph and Graph are as defined in §1.

The coherence theorem in the preceding section establishes conditions under which two arrows of Shap (the category of shapes of a formal closed category V) are equal when "translate" in the concrete elements of the category V: the result is handy to use because the computation of the graph of the arrows of Shap can be done by means of almost mechanical drawings. Any closed category is formal closed so that our results apply immediately to closed categories.

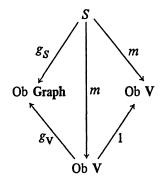
We are going to complete our work by proving that the graph determines the arrows of F(S) within equivalences.

PROPOSITION 11. Let $f, g: X \to Y$ be two arrows of F(S) with the same domain and codomain. Then f and g are equivalent if and only if graph(f) = graph(g).

PROOF. The definitions of graph and equivalence imply that if f and g are equivalent then graph(f) = graph(g) and this proves a part of the proposition.

Suppose that V is a formal closed category, Shap the category of shapes of V, $m: S \to Ob V$ an *I*-preserving map and $m': F(S) \to V$ the functor obtained by means of the adjunction.

Note that we have a commutative diagram



where the maps g_S and g_V have been defined in §1. We have to remark also that the adjunction takes the maps g_S , g_V and 1 onto the functors

graph_S:
$$F(S) \rightarrow$$
 Graph,
graph_V: Shap \rightarrow Graph,
 $| : Shap \rightarrow V,$

respectively.

The commutativity of the above diagrams implies the relations,

$$m' = | | \cdot F(m), \quad \operatorname{graph}_{S} = \operatorname{graph}_{V} \circ F(m),$$

and we have the chain of implications,

$$\operatorname{graph}_{S}(f) = \operatorname{graph}_{S}(g)$$

 $\Rightarrow \operatorname{graph}_{V}(F(m)(f)) = \operatorname{graph}_{V}(F(m)(g))$
 $\Rightarrow |F(m)(f)| = |F(m)(g)| \Rightarrow m'(f) = m'(g).$

Thus $graph_S(f) = graph_S(g)$ implies that f and g are equivalent and we have proved the proposition.

We can now give a construction that clarifies the meaning of the equivalence of arrows in F(S). If we consider the quotient of F(S) by the equivalence of arrows we obtain a category F'(S). It is very easy to check that F'(S) is a formal closed category that has the following universal property: if V is a formal closed category and $m: S \to \text{Ob V}$ is a map such that m(I) = I then there is exactly one functor $m': F'(S) \to V$ that preserves the structure and reduces to m over $S \subset \text{Ob } F'(S)$. In other words, F' is the left adjoint of the forgetful functor G' from the category of formal closed categories and strict-structure-preserving functors onto the sets with fixed point I: G' takes each category onto its objects. Thus the solution of the coherence problem gives a useful tool to describe the free categories F'(S): one could also take the viewpoint of considering the description of F'(S) as the relevant coherence problem [10, p. 283].

Moreover F'(S) is a closed category and this proves that F' is also the left adjoint for the restriction of G' to the full subcategory of closed categories. This fact seems to be equivalent to an opinion of G. M. Kelly [9, p. 52] and it is another confirmation that the coherence problem in closed and formal closed categories is essentially the same.

REFERENCES

- 1. B. Day, An embedding theorem for closed categories, Category Seminar (Proc. Sem., Sidney, 1972/73), Lecture Notes in Math., vol. 420, Springer, Berlin, 1974, pp. 55-64. MR 51 #650.
- 2. B. Day and M. Laplaza, An embedding theorem for closed categories, Abstracts of the Sydney Category Theory Seminar 1972, The University of South Wales, pp. 53-56.

- 3. S. Eilenberg and G. M. Kelly, *Closed categories*, Proc. Conf. Categorical Algebra (La Jolla, Calif., 1965), Springer, New York, 1966, pp. 421-562. MR 37 #1432.
- 4. _____, A generalization of the functorial calculus, J. Algebra 3 (1966), 366-375. MR 32 #7618.
- 5. G. Gentzen, Untersuchungen über das logische Schliessen I, II, Math. Z. 39 (1934/35), 176-210; 405-431.
 - 6. J. R. Isbell, Review of [10]. MR 44 #278.
- 7. G. M. Kelly, *Many-variable functorial calculus*. I, Coherence in Categories, Lecture Notes in Math., vol. 281, Springer, Berlin, 1972, pp. 66-105. MR 49 #5126.
- 8. _____, An abstract approach to coherence, Coherence in Categories, Lecture Notes in Math., vol. 281, Springer, Berlin, 1972, pp. 106-147. MR 49 #5127.
- 9. _____, Mixed-variance clubs, Abstracts of the Sydney Category Theory Seminar 1972, The University of South Wales, pp. 46-53.
- 10. _____, Coherence theorems for Lax algebras and for distributive laws, Category Seminar (Proc. Sem., Sidney, 1972/73), Lecture Notes in Math., vol. 420, Springer, Berlin, 1974, pp. 281-375. MR 51 #648.
- 11. G. M. Kelly and S. Mac Lane, Coherence in closed categories, J. Pure Appl. Algebra 1 (1971), no. 1, 97-140; erratum, ibid. 1 (1971), no. 2, 219. MR 44 #278; 45 #1988.
- 12. _____, Closed coherence for a natural transformation, Coherence in Categories, Lecture Notes in Math., vol. 281, Springer, Berlin, 1972, pp. 1-28. MR 51 #10437.
- 13. J. Lambek, Deductive systems and categories. I: Syntactic calculus and residuated categories, Math. Systems Theory 2 (1968), 287-318. MR 38 #4277.
- 14. _____, Deductive systems and categories. II: Standard constructions and closed categories, Category Theory, Homology Theory and Their Applications, I (Battelle Inst. Conf., Seattle, Wash., 1968, Vol. One), Lecture Notes in Math., vol. 86, Springer-Verlag, New York, 1969, pp. 76–122. MR 39 #3967.
- 15. _____, Deductive systems and categories, III: Cartesian closed categories, intuitionist propositional calculus and combinatory logic, Toposes, Algebraic Geometry and Logic (Conf., Dalhousie Univ., Halifax, N.S., 1971), Lecture Notes in Math., vol. 274, Springer, Berlin, 1972, pp. 57–82. MR 50 #1850.
- 16. M. Laplaza, Coherence for associativity not an isomorphism, J. Pure Appl. Algebra 2 (1972), no. 2, 107-120. MR 45 #6885.
- 17. _____, Coherence for categories with associativity, commutativity and distributivity, Bull. Amer. Math. Soc. 78 (1972), 220-222. MR 45 #5193.
- 18. _____, Coherence for distributivity, Coherence in Categories, Lecture Notes in Math., vol. 281, Springer, Berlin, 1972, pp. 29-65. MR 49 #379a.
- 19. _____, Embedding of closed categories into monoidal closed categories, Trans. Amer. Math. Soc. (to appear).
- 20. G. Lewis, Coherence for a closed functor, Coherence in Categories, Lecture Notes in Math., vol. 281, Springer, Berlin, 1972, pp. 148–195. MR 49 #9048.
- 21. S. Mac Lane, Natural associativity and commutativity, Rice Univ. Studies 49 (1963), no. 4, 28-46. MR 30 #1160.
 - 22. _____, Categorical algebra, Bull. Amer. Math. Soc. 71 (1965), 40-106. MR 30 #2053.
- 23. _____, Coherence and canonical maps, Symposia Mathematica, Vol. IV (INDAM, Rome, 1968/69), Academic Press, London, 1970, pp. 231-242. MR 46 # 1879.
- 24. _____, Categories for the working mathematician, Springer-Verlag, Berlin and New York, 1971. MR 50 #7275.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF PUERTO RICO, MAYAGUEZ, PUERTO RICO 00708

Current address: 14 M-7-B Terrace, Mayaguez, Puerto Rico 00708