Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Pure and Applied Algebra 198 (2005) 57–65
JOURNAL OF PURE AND APPLIED ALGEBRA
www.elsevier.com/locate/jpaa

Coherence of associativity in categories with multiplication

Matthew G. Brin

Department of Mathematical Sciences, State University of New York at Binghamton, Binghamton, NY 13902-6000, USA

Received 13 August 2004
Available online 24 November 2004
Communicated by C.A. Weibel

Abstract

We show that any associativity isomorphism in a category with multiplication is coherent in the sense of MacLane if the operations for building new isomorphisms from it are restricted so that tensoring with the identity is only allowed on the right instead of on both the right and the left. With this restriction, coherence is obtained without the assumption that the pentagon diagram commutes.

MSC: Primary: 18D10; secondary: 20F05

1. Introduction

To say that $\mathscr{C}$ is a category with (functoral) multiplication means that there is a functor $\otimes : \mathscr{C}^2 \to \mathscr{C}$ called the multiplication where $\mathscr{C}^2$ is the category of pairs of objects and pairs of morphisms from $\mathscr{C}$. [More technically, $\mathscr{C}^2$ is the category of functors and natural transformations from 2 to $\mathscr{C}$ where 2 is the category with objects 0 and 1, and the only morphisms are the identity morphisms.] Examples of functoral multiplications are cross products, tensor products, free products and so forth on those categories where those products exist.

For most examples it is rarely the case that

$$A \otimes (B \otimes C) = (A \otimes B) \otimes C \tag{1}$$

E-mail address: matt@math.binghamton.edu (M.G. Brin).

0022-4049/$ - see front matter
doi:10.1016/j.jpaa.2004.10.008

is literally true, and what is usually the case is that there is a natural isomorphism α from the functor $F : \mathscr{C}^3 \to \mathscr{C}$ defined by $F(A, B, C) = A \otimes (B \otimes C)$ to the functor $G : \mathscr{C}^3 \to \mathscr{C}$ defined by $G(A, B, C) = (A \otimes B) \otimes C$. In the most common cases, there is an obvious candidate for natural isomorphism α and it is a triviality to define.

The usual statement that "all associativity laws follow from the associativity law given in (1)" translates into a claim that if H and K are two functors from $\mathscr{C}^n$ to $\mathscr{C}$ that are built by combining n variables in the same order with $n - 1$ applications of $\otimes$ and that differ only in the pattern of parentheses, then there is a natural isomorphism from H to K that is derivable in some sensible way from α. The problem might be that there is more than one way to build such an isomorphism from α, raising the possibility that different ways will result in different isomorphisms.

This problem was first considered by MacLane in [6], where he defined the condition *coherence* of such an α to mean that any two expressions built from $\otimes$ using the same variables in the same order and differing only in the distribution of parentheses are connected by a unique natural isomorphism derivable from α using a prescribed set of constructions. In [6] it is proven that coherence is achieved from the naturality of α and one hypothesis that a certain (now famous) pentagonal diagram commutes.

The purpose of this paper is to show that the hypothesis that the pentagonal diagram commute can be dispensed with if the prescribed set of constructions for building natural isomorphisms from α is restricted. Thus, we do not prove a strengthening or generalization of MacLane's theorem. It is simply a different theorem.

Beyond the statement and proof of this theorem, the paper has a second purpose which is to point out the connection between MacLane's theorem on coherence and combinatorial group theory. This is discussed in the last section, where we point out that MacLane's theorem can be viewed as giving a presentation of a certain group in terms of generators and relations.

2. Statement

The constructions in [6] for building isomorphisms from α are extremely natural. (Overuse of the word *natural* here is unavoidable.) The restrictions on the constructions in this paper lack a certain symmetry. Thus, our result suffers from a certain aesthetic inferiority. We now give some details and start with some preliminary technicalities.

If β is a natural transformation from a functor $F : \mathscr{A} \to \mathscr{B}$ to a functor $G : \mathscr{A} \to \mathscr{B}$, then we can view β as a functor from $\mathscr{A}$ to $\mathscr{B}^{\overline{2}}$, the category of functors from $\overline{2}$ to $\mathscr{B}$ in which $\overline{2}$ is the category with objects 0 and 1 and only one non-identity morphism that goes from 0 to 1. The category $\overline{2}$ is just the category whose objects are 0 and 1 and whose morphisms correspond to the partial order $\leqslant$; while $\mathscr{B}^{\overline{2}}$ is just the category whose objects are the morphisms of $\mathscr{B}$ and whose morphisms are the commutative squares in $\mathscr{B}$. If S is the "source" functor from $\mathscr{B}^{\overline{2}}$ to $\mathscr{B}$ in which $S(f : X \to Y) = X$ and T is the "target" functor in which $T(f : X \to Y) = Y$, then $S\beta = F$ and $T\beta = G$.

Any functor $F : \mathscr{A} \to \mathscr{B}$ induces a functor $F^{\overline{2}} : \mathscr{A}^{\overline{2}} \to \mathscr{B}^{\overline{2}}$.

In [6] isomorphisms are built from $\alpha : A \otimes (B \otimes C) \to (A \otimes B) \otimes C$ by four processes. The one that we will restrict is as follows.

If β is a natural transformation from functor F to functor G that each go from $\mathscr{C}^m$ to $\mathscr{C}$ and γ is a natural transformation from H to K that each go from $\mathscr{C}^n$ to $\mathscr{C}$, then we can form $\beta \otimes \gamma$ going from $F \otimes H$ to $G \otimes K$ by composing

$$\beta \times \gamma : \mathscr{C}^m \times \mathscr{C}^n \to \mathscr{C}^{\overline{2}} \times \mathscr{C}^{\overline{2}}$$

with

$$\otimes^{\overline{2}} : \mathscr{C}^{\overline{2}} \times \mathscr{C}^{\overline{2}} \to \mathscr{C}^{\overline{2}}.$$

The operation $\otimes$ on transformations can be used for the following. Let $\mathbf{1}$ denote the identity transformation from the identity functor on $\mathscr{C}$ to itself. We can then form $\alpha \otimes \mathbf{1}$, $(\alpha \otimes \mathbf{1}) \otimes \mathbf{1}$ and so forth where, inductively, $\alpha_0 = \alpha$ and $\alpha_i = \alpha_{i-1} \otimes \mathbf{1}$. Thus

$$\alpha_1 = \alpha \otimes \mathbf{1} : (A \otimes (B \otimes C)) \otimes D \to ((A \otimes B) \otimes C) \otimes D$$

with similar descriptions of other α_i. We can refer to $\beta \otimes \mathbf{1}$ as the *right stabilization* of β. We refer to the α_i as the iterated right stabilizations of α.

The assumptions in [6] are that the transformations form a category closed (among other things) under the operation $\otimes$ on transformations. In this paper, we will only make use of the operation $\otimes$ on transformations to create right stabilizations. All other constructions from [6] will be used here. We now go on to the others.

From

$$\alpha : A \otimes (B \otimes C) \to (A \otimes B) \otimes C \tag{2}$$

we can create

$$\alpha' : (A \otimes B) \otimes (C \otimes D) \to ((A \otimes B) \otimes C) \otimes D$$

from (2) by replacing A in (2) by the product of two variables. Similarly, we get

$$\alpha'' : A \otimes ((B \otimes C) \otimes D) \to (A \otimes (B \otimes C)) \otimes D$$

by replacing B in (2) by the product of two variables. These are both examples of *instances* of α. More generally, we can replace any variable in (2) on both sides by identical expressions involving $\otimes$.

Technically, an *instance* of a transformation is created by precomposing the transformation with a functor. Now if $\beta : \mathscr{A} \to \mathscr{B}^{\overline{2}}$ is a natural transformation from $F = S\beta$ to $G = T\beta$, and if $H : \mathscr{D} \to \mathscr{A}$ is any functor, then βD is a natural transformation from $FD = S\beta D$ to $GD = T\beta D$ and can be viewed as an instance of β.

In our setting, we will take instances of the iterated right stabilizations α_i of α. The iterated stabilization of α_i connects functors from $\mathscr{C}^n$ to $\mathscr{C}$ where $n = i + 3$. Instances can be created by precomposing the stabilizations with compositions of functors such as

$$(X_1, X_2, \ldots, X_j, X_{j+1}, \ldots, X_{m+1}) \mapsto (X_1, X_2, \ldots, X_j \otimes X_{j+1}, \ldots, X_{m+1})$$

from $\mathscr{C}^{m+1}$ to $\mathscr{C}^m$ for various values of m and j.

We will also postcompose a transformation with a functor. If $\beta : \mathscr{A} \to \mathscr{B}^{\bar{2}}$ is a natural transformation and $J : \mathscr{B} \to \mathscr{E}$ is a functor, then $J\beta$ represents the composition of β

with $J^{\bar{2}}$. This construction can yield an instance (of another transformation) by accident, and we will exploit this.

Another operation for constructing isomorphisms from α is that of composition. If F, G and H are all functors from $\mathscr{A}$ to $\mathscr{B}$, if β is a natural transformation from F to G and γ is a natural transformation from G to H, then there is an obvious composition $\gamma\beta$ that is a natural transformation from F to H. Composition commutes with right stabilization.

The final operation for constructing isomorphisms from α is that of inversion. Since α, its stabilizations and its instances are all isomorphisms, they are all invertible. Note that inversion commutes with instance and stabilization and behaves in the usual way with respect to composition: $(\beta\gamma)^{-1} = \gamma^{-1}\beta^{-1}$.

We can now state our result.

Theorem 1. *Let $\mathscr{C}$ be a category with functoral multiplication $\otimes : \mathscr{C}^2 \to \mathscr{C}$. Let α be a natural isomorphism from $A \otimes (B \otimes C)$ to $(A \otimes B) \otimes C$. If E and F are two expressions in $n - 1$ appearances of $\otimes$ and n different variables in the same order that differ only in the arrangement of parentheses, then there exists a unique natural isomorphism constructable from α as a composition of instances of iterations of right stabilizations of α and α^{-1}.*

3. Proof

The proof of Theorem 1 is essentially the proof of Theorem 3.1 of [6] with more attention paid to some details. We will include the entire proof since a set of instructions on modifying the proof in [6] would be unreadable.

We will discuss expressions endlessly. For us an *expression* in n variables is a fully parenthesized alternation of the variables $X_1, \ldots, X_n$ in that order with $n - 1$ appearances of the operation $\otimes$. Inductively, the variable X_1 is the only expression in 1 variable, and if F and G are expressions in m and n variables, respectively, then $(F \otimes \overline{G})$ is an expression in $m + n$ variables where $\overline{G}$ is the expression G with all the subscripts of its variables raised uniformly by m. We will omit the bar from the second expression from now on since the meaning will always be clear.

An expression is *trivial* if it has only one variable. We reserve the symbol $\mathbf{I}$ to symbolize the trivial expression X_1.

A non-trivial expression E breaks uniquely as $(F \otimes G)$. We say that E is *semi-normalized* if $E = (F \otimes \mathbf{I})$. We can refer to $(F \otimes \mathbf{I})$ as the right stabilization of F. Right stabilization can be iterated and we define $(F \bigotimes_i \mathbf{I})$ inductively by $(F \bigotimes_0 \mathbf{I}) = F$ and $(F \bigotimes_i \mathbf{I}) = ((F \bigotimes_{i-1} \mathbf{I}) \otimes \mathbf{I})$.

An expression in n variables is *fully normalized* if it is of the form $(\mathbf{I} \bigotimes_{n-1} \mathbf{I})$. There is only one fully normalized expression on n variables for each n and we will denote it by $\mathbf{I}_n$. We have

$$\mathbf{I}_n = (\cdots(((\mathbf{I} \otimes \mathbf{I}) \otimes \mathbf{I}) \otimes \mathbf{I}) \otimes \cdots \otimes \mathbf{I}), \quad (n \text{ appearances of } \mathbf{I}).$$

If an expression E is not fully normalized, then it is uniquely expressible as $(N \bigotimes_i \mathbf{I})$ where N is not semi-normalized. If E is not semi-normalized, then $i = 0$. The value of i is the *normalization level* of E. Note further that $N = (F \otimes G)$ for some F and G with $G \neq \mathbf{I}$. The *weight* of E is the number of variables used in G.

If an expression E on n variables is not fully normalized, then its normalization level is strictly less than n and its weight is strictly greater than 1. We extend the definitions to say that the normalization level of $\mathbf{I}_n$ is n and that its weight is 1. (There is only one expression on two variables and it is fully normalized, so the normalization level of an n variable expression with $n \geqslant 2$ is never $n-1$.)

The point of all this book keeping is the list of observations below. They are verified by inspecting the form of the various α_i. We say that a natural transformation from a functor F to a functor G has F as its source and G as its target. We are treating expressions formally, but they represent functors. Thus, we can talk about instances of the α_i as having expressions for source and target. The number of variables of the source and target of a given instance of an α_i will be the same. In reading the following, note that we carefully distinguish between α_i and α_i^{-1} and the fact that α_i^{-1} is never mentioned is significant.

(A1) If E is an n variable expression, then for each i there is at most one instance of α_i that can have E as source.
(A2) If E is an n variable expression, then an instance of α_i can have E as source only if $i \leqslant n-3$.
(A3) If E is an n variable expression with normalization level k, then an instance of α_i can have E as source only if $i \geqslant k$.
(A4) If E is an n variable expression with normalization level k and weight $w > 1$, then an instance of α_i having E as source with $i > k$ has a target with normalization level k and weight w.
(A5) If E is an n variable expression with normalization level k and weight $w > 1$, then there is an instance of α_k having E as source. Further the target of this instance of α_k either has normalization level that is greater than k or has normalization level equal to k and weight less than w.

If E is an expression and a string $\alpha_{i_1}\alpha_{i_2}\cdots\alpha_{i_s}$ has the property that an instance of α_{i_s} has E as a source and target F_s, and for each $j < s$ an instance of α_{i_j} has F_{j+1} as a source and target F_j, then we say that the string is a word in the α_i that defines a path from E to F_1. Note that the information in the string does not specify which instances are used, but this is not necessary because of (A1).

If an instance of α_i has source E and target F, then the instance is a natural isomorphism from the functor represented by E to the functor represented by F. Thus in the previous paragraph, the word in the α_i defines an isomorphism from E to F_1.

It is now an easy inductive exercise to prove the following from (A1)–(A5).

Lemma 1. *Given an expression E in n variables that is not fully normalized, then there is a unique word*

$$w = \alpha_{i_1}\alpha_{i_2}\cdots\alpha_{i_s}$$

satisfying $i_j \geqslant i_k$ if $j < k$ so that w is an isomorphism from E to $\mathbf{I}_n$.

This proves the existence part of Theorem 1 since any two expressions in n variables can be connected to $\mathbf{I}_n$ by an isomorphism.

We now continue with our reading of the proof from [6]. If F and G are two expressions in n variables, then we must show that any two "paths" from F to G where each step in each path is an instance of some α_i or α_i^{-1}, then the two paths compose to the same natural isomorphism from F to G.

Our argument begins as it does in [6]. We direct each step in the path by declaring that each step in the path goes from the source expression of the instance of α_i to its target expression. Take an arbitrary path from F to G, and join each vertex F_i in the path to $\mathbf{I}_n$ by the path obtained from Lemma 1. Note that this path is directed from F_i to $\mathbf{I}_n$. This creates a diagram of which the following sample is typical.

$$\begin{array}{ccccccccc} F & \longrightarrow & F_1 & \longleftarrow & F_2 & \longrightarrow & F_3 & \longleftarrow & G \\ {\scriptstyle p}\downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow{\scriptstyle q} \\ I_n & = & I_n & = & I_n & = & I_n & = & I_n \end{array}$$

If it is shown that the above diagram commutes, then the path along the top from F to G gives the same isomorphism as $q^{-1}p$ and the proof of Theorem 1 will be complete. Thus it suffices to prove the commutativity of a single rectangle of the form

$$\begin{array}{ccc} F & \xrightarrow{\alpha_j} & G \\ {\scriptstyle p}\downarrow & & \downarrow{\scriptstyle q} \\ I_n & = & I_n \end{array} \tag{3}$$

in which the top arrow is an instance of some α_j and the paths p and q are obtained from Lemma 1.

The expression F has a normalization level i so we know that the first step in p (the last letter expressing p as a word) is α_i and we know $j \geqslant i$. If $j = i$, then the uniqueness gotten from Lemma 1 says that p and $q\alpha_j$ are identical as words and the rectangle (3) commutes. Thus we are left with the case $j > i$.

If $j > i$, then the normalization level of G is also i by (A4) and q also has α_i as its first step. Thus, we will be done by induction on the length of p when we show that the following rectangle commutes whenever $j > i$.

$$\begin{array}{ccc} F & \xrightarrow{\alpha_j} & G \\ {\scriptstyle \alpha_i}\downarrow & & \downarrow{\scriptstyle \alpha_i} \\ F_1 & \xrightarrow{\alpha_j+1} & G_1 \end{array} \tag{4}$$

The expression F equals $(N \bigotimes_i \mathbf{I})$ with $N = (H \otimes (K \otimes L))$, and $F_1 = (N' \bigotimes_i \mathbf{I})$ with $N' = ((H \otimes K) \otimes L)$. Since $j > i$, we know that $G = (N'' \bigotimes_i \mathbf{I})$, where $N'' = (H' \otimes (K \otimes L))$ is the result of applying α_{j-i} to N and so H' is the result of applying α_{j-i-1} to H. Expanding what we know about F_1 and applying α_i to G gives

$$F_1 = (((H \otimes K) \otimes L) \bigotimes_i \mathbf{I})$$

and

$$G_1 = (((H' \otimes K) \otimes L)\bigotimes_i \mathbf{I}).$$

This says that G_1 is the result of applying α_{j+1} to F_1.

This does not make (4) commute. It only says that source and targets make sense. That commutativity follows from the naturality of α_i can be seen by filling in the details of (4) to give the following.

$$\begin{tikzcd}
((H \otimes (K \otimes L))\bigotimes_i \mathbf{I}) \arrow[r, "\alpha_j"] \arrow[d, "\alpha_i"'] & ((H' \otimes (K \otimes L))\bigotimes_i \mathbf{I}) \arrow[d, "\alpha_i"'] \\
(((H \otimes K) \otimes L)\bigotimes_i \mathbf{I}) \arrow[r, "\alpha_{j+1}"] & (((H' \otimes K) \otimes L)\bigotimes_i \mathbf{I})
\end{tikzcd} \tag{5}$$

If we define functors R and L by

$$R(-) = ((- \otimes (K \otimes L))\bigotimes_i \mathbf{I})$$

and

$$L(-) = (((- \otimes K) \otimes L)\bigotimes_i \mathbf{I}),$$

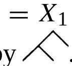
then the specific instances of α_j and α_{j+1} in (5) are seen to be $R\alpha_{j-i-1}$ and $L\alpha_{i-j-1}$, respectively. Further, both appearances of α_i are instances of a single natural isomorphism $\overline{\alpha}_i$ with source R and target L, and $\overline{\alpha}_i$ is an instance of α_i. Diagram (5) commutes since the following diagram commutes by the naturality of $\overline{\alpha}_i$.

$$\begin{tikzcd}
RH \arrow[r, "R\alpha_{i-j-1}"] \arrow[d, "\overline{\alpha}_i"'] & RH' \arrow[d, "\overline{\alpha}_i"] \\
LH \arrow[r, "L\alpha_{i-j-1}"] & LH'
\end{tikzcd}$$

This completes the proof of Theorem 1. □

4. The origins of Theorem 1 and its Proof

Theorem 1 is a thinly disguised translation of the well-known fact that a certain group has a certain presentation. The proof that we give of Theorem 1 contains much of the work from the standard (and well-known) proofs of this well-known fact.

There is a group commonly known as Thompson's group F (see [2]) that is possessed of many descriptions. One description in [2] uses pairs of finite binary trees. Parenthesized expressions are captured by trees. Thus $E_1 = X_1 \otimes (X_2 \otimes X_3)$ is captured by the tree  and $E_2 = (X_1 \otimes X_2) \otimes X_3$ is captured by . We can summarize the fact that α_0 has E_1 as source and E_2 as target by writing

$\alpha_0 = \Big($  ,  $\Big)$.

Our isomorphisms connect paired expressions having the same number of variables, and these paired expressions correspond to pairs of trees that have the same number of leaves. The elements of the group F are the equivalence classes of all pairs of finite binary trees in which the two trees in the pair have the same number of leaves. It is easiest to explain the equivalence relation put on such pairs of trees by saying that two pairs are equivalent if they correspond to instances of the same isomorphism. The multiplication of pairs is defined by writing $(T_1, T_2)(T_2, T_3) = (T_1, T_3)$. This multiplies elements in the reverse order that we have composed isomorphisms, so the discussion that follows will have some flips in it. The arguments that all equivalence classes can be multiplied in a well-defined manner and that the resulting multiplication gives a group can be found in [2].

Those familiar with Thompson's group F will recognize the proof of Theorem 1 as the bulk of yet another proof that F has a certain presentation. With our right-to-left convention for composing isomorphisms, we end up with the non-standard version of the presentation that reads

$$F = \langle \alpha_0, \alpha_1, \ldots \mid \alpha_i \alpha_j = \alpha_{j+1} \alpha_i, \ \text{whenever } i < j \rangle.$$

The usual presentation would have the relations read $\alpha_j \alpha_i = \alpha_i \alpha_{j+1}$ when $i < j$.

That Thompson's group F is closely associated to associativity is well known. See [3] and the end comments of [7]. Further, given a category $\mathscr{C}$ with multiplication $\otimes$ and associativity isomorphism α, it is possible to define a group $G(\mathscr{C}, \otimes, \alpha)$ that will be isomorphic to F if and only if $(\mathscr{C}, \otimes, \alpha)$ is coherent in the sense of [6] (and not in our more restrictive sense). This statement is nothing more than checking of definitions. There is a similar statement connecting the symmetric, monoidal categories (which combine associativity and commutativity) with another of Thompson's groups known as V. Again, this is just a check of definitions and repeats a well-known connection between V and the pair consisting of associativity and commutativity (see [4]). There is a less trivial connection between the braided tensor categories of [5] and a braided version of V constructed in [1,4]. This connection will be explored elsewhere.

The usual theorem on coherence of associativity, Theorem 3.1 of [6], involves the full power of the operation $\otimes$ on natural isomorphisms. Although not apparent in [6], the main effect of this is to introduce the left stabilizations of α. A glance at the pentagon diagram (3.5) of [6] shows that the diagram can be used to express the left stabilization $\mathbf{1} \otimes \alpha$ in terms of the right stabilization $\alpha \otimes \mathbf{1}$ and instances of α.

As a final remark, we point out that F has a presentation with only the generators α_0 and α_1 and only two relations. That α_0 and α_1 suffice to generate follows from the relations given and also from the commutativity of the diagram (4). If desired, Theorem 1 can be restated to end with the words: ... *unique natural isomorphism constructed as a composition of instances of* α_0, α_1 *and their inverses*. There is nothing to be learned from the small number of relations since naturality gives all the relations that are needed and more.

References

[1] M.G. Brin, The algebra of strand splitting. I. The algebraic structure of the braided Thompson group, preprint, Binghamton University, 2004.

[2] J.W. Cannon, W.J. Floyd, W.R. Parry, Introductory notes on Richard Thompson's groups, Enseign. Math. (2) 42 (3–4) (1996) 215–256 MR 98g:20058.

[3] P. Dehornoy, The structure group for the associativity identity, J. Pure Appl. Algebra 111 (1–3) (1996) 59–82 MR 97d:55033.

[4] P. Dehornoy, Geometric presentations for Thompson's groups and related groups, preprint, University of Caen, 2004.

[5] A. Joyal, R. Street, Braided tensor categories, Adv. Math. 102 (1) (1993) 20–78 MR 94m:18008.

[6] S. MacLane, Natural associativity and commutativity, Rice Univ. Studies 49 (4) (1963) 28–46.

[7] R. McKenzie, R.J. Thompson, An elementary construction of unsolvable word problems in group theory, in: W.W. Boone, F.B. Cannonito, R.C. Lyndon (Eds.), Word Problems, North-Holland, Amsterdam, 1973, pp. 457–478.