1 COMBINATORIAL GROUP THEORY

1.1 *FREE GROUPS*

Let X be a generating subset of a group G. Certain products of members of X and their inverses will be 1 whatever X and G are; for instance, $xyyz^{-1}zy^{-1}y^{-1}x^{-1}$. Other products, such as xyz or xx, will be 1 for some choices of X and G but not for other choices. Those pairs G and X for which a product of elements in $X \cup X^{-1}$ is 1 only when the properties holding in all groups require it to be 1 are obviously of interest.

They are called *free groups;* a more formal definition will be given later. If G is such a group, any function f from X to a group H can be extended uniquely to a homomorphism from G to H. For any $g \in G$ can be written as $x_{i_1}^{\varepsilon_1} \dots x_{i_n}^{\varepsilon_n}$ where $\varepsilon_r = \pm 1$ and $x_{i_r} \in X$ for $r = 1, \dots, n$. Now suppose that g can also be written as $x_{j_1}^{\delta_1} \dots x_{j_m}^{\delta_m}$, where $\delta_s = \pm 1$ and $x_{j_s} \in X$ for $s = 1, \dots, m$. Then

$$x_{i_1}^{\varepsilon_1} \dots x_{i_n}^{\varepsilon_n} x_{j_m}^{-\delta_m} \dots x_{j_1}^{-\delta_1} = 1$$

and our assumption on G and X then tells us we must have

$$(x_{i_1}f)^{\varepsilon_1} \dots (x_{i_n}f)^{\varepsilon_n} (x_{j_m}f)^{-\delta_m} \dots (x_{j_1}f)^{-\delta_1} = 1.$$

Hence the element of H given by $(x_{i_1}f)^{\varepsilon_1} \dots (x_{i_n}f)^{\varepsilon_n}$ depends only on g and not on how g is written as a product of elements of $X \cup X^{-1}$. It follows that we can define a function $\varphi: G \to H$ by requiring $g\varphi$ to be this element. It is easy to check that φ is a homomorphism and that $x\varphi = xf$ for all $x \in X$. Since every element of G is a product of elements of X and their inverses, there can only be one homomorphism with specified values on X. Conversely, suppose that X is a subset of a group G such that any function from X into an arbitrary group H extends uniquely to a homomorphism from G to H. Let $x_{i_1}^{\varepsilon_1} \dots x_{i_n}^{\varepsilon_n}$ equal 1 in G, and let f be a function from X into a group H. Since f extends to a homomorphism from G to H, it follows that $(x_{i_1}f)^{\varepsilon_1} \dots (x_{i_n}f)^{\varepsilon_n}$ equals 1 in H.

Since this holds for all H and f, this amounts to saying that the product of elements of $X \cup X^{-1}$ equals 1 in G only if the corresponding products in all groups are also 1. It can be shown that this holds iff we can derive the fact that the product equals 1 from the group laws. These observations lead to the formal definition of free groups.

Definition Let X be a set, G a group, and $i: X \to G$ a function. The pair (G,i) is called *free on* X if for every group H and function $f: X \to H$ there is a unique homomorphism $\varphi: G \to H$ such that $f = i\varphi$.

In particular, the trivial group is free on the empty set, and the infinite cyclic group $\mathbb{Z}$ is free on any one element set $\{x\}$, where xf is the integer 1.

The following three questions are natural ones to ask. Do free groups on an arbitrary set exist? Are free groups on a given set unique? If (G,i) is free on X is i injective? We leave the first question till later. The answer to the second question and a partial answer to the third are of an abstract nature, applying in many situations in abstract algebra and category theory.

It will be useful at times to use the language of category theory, which frequently sheds light on general situations. (Barry Mitchell once wrote "The purpose of category theory is to show that what is trivial is trivially trivial.") For instance, in asking whether free groups on an arbitrary set exist we are asking whether the forgetful functor from groups to sets has an adjoint, and the fact that free groups on a given set are essentially unique is just a uniqueness property of adjoints (or of initial objects in categories). Readers who have no knowledge of category theory can safely ignore any remarks about it; such remarks are made only to give extra insight to those with the relevant knowledge.

Plainly, if (G,i) is free on X and $\varphi: G \to H$ is an isomorphism then $(H,i\varphi)$ is also free on X. Our first proposition is the converse of this.

Proposition 1 *Let* (G_1,i_1) *and* (G_2,i_2) *be free on* X. *Then there is an isomorphism* $\varphi: G_1 \to G_2$ *such that* $i_1\varphi = i_2$.

Proof Since (G_1,i_1) is free on X, there is, by definition, a homomorphism $\varphi: G_1 \to G_2$ such that $i_1\varphi = i_2$. Similarly, there is a homomorphism $\psi: G_2 \to G_1$ such that $i_2\psi = i_1$. We then have $i_1\varphi\psi = i_1 = i_1 I_1$, where I_1 is the identity function on G_1. The uniqueness property in the definition now requires that

$\varphi\psi = I_1$, and, similarly, $\psi\varphi = I_2$, the identity function on G_2. Hence φ is an isomorphism.//

Proposition 2 *Let (F,i) be free on X.*

(i) If there is a group G with an injective function from X to G then i is injective.

(ii) There is a group into which X maps injectively; for instance, the set $\mathbb{Z}^X$ of all functions from X to $\mathbb{Z}$, where $\alpha+\beta$ is defined by $x(\alpha+\beta) = x\alpha + x\beta$ for all x.

(iii) i is injective.

Proof (i) Let $f: X \to G$ be an injection. Then there is a homomorphism $\varphi: F \to G$ such that $f = i\varphi$. It then follows that i is injective. Notice that this part of the argument is of an abstract nature, holding in many situations, whereas (ii) requires a specific example.

(ii) Plainly $\mathbb{Z}^X$ is a group. If we define α_x by $x\alpha_x = 1$ and $y\alpha_x = 0$ for $y \neq x$ then the function sending x to α_x is plainly injective. Part (iii) is now immediate.//

We now proceed to construct a group free on X, beginning with an auxiliary construction. Let X be any set, and let $M(X)$ denote the set of all finite sequences $(x_{i_1}, \ldots, x_{i_n})$ of elements of X, where $n \geq 0$ (the case $n = 0$ corresponds to the empty sequence). Define a multiplication on $M(X)$ by

$$(x_{i_1}, \ldots, x_{i_n})(x_{j_1}, \ldots, x_{j_m}) = (x_{i_1}, \ldots, x_{i_n}, x_{j_1}, \ldots, x_{j_m}).$$

This multiplication is obviously associative, with an identity element which we call 1 (namely, the empty sequence). Also, $x \to (x)$ is obviously one-one, and, if we identify x with (x), every element of $M(X)$ can be uniquely written as a product $x_{i_1} \ldots x_{i_n}$ for some n; we shall always use this notation. We call $M(X)$ the *free monoid* on X.

By a *segment* of $x_{i_1} \ldots x_{i_n}$ we mean an element $x_{i_r} x_{i_{r+1}} \ldots x_{i_s}$, where $1 \leq r \leq s \leq n$; this is called an *initial segment* if $r = 1$, and a *final segment* (or *terminal segment*) if $s = n$, and it is a *proper segment* unless $r = 1$ and $s = n$.

One technical point has been glossed over in this construction. Since X can be any set, it is possible that some element of X is itself a finite sequence of other elements of X. We would then want to distinguish between this sequence as an element of X and as an element of $M(X)$. The simplest way of dealing with this problem is to replace X by the set X' which is defined to

be the set of all $\{x\}$ for $x \in X$, and to define the free monoid on X to be $M(X')$. However, we shall ignore this technicality in future.

We now proceed to construct the free group on X. Take a set $\bar{X}$ bijective with X under a bijection which sends x to $\bar{x}$, and such that $X \cap \bar{X} = \emptyset$. We usually denote $\bar{x}$ by x^{-1}, and we may also write x^1 instead of x. The elements of $M(X \cup \bar{X})$ are called *words on* X. Let w be the word $x_{i_1}^{\varepsilon_1} \dots x_{i_n}^{\varepsilon_n}$. Then n is called the *length* of w, written $|w|$ or $l(w)$, and we call the elements $x_{i_r}^{\varepsilon_r}$ the *letters* of w.

The word w is called *reduced* if, for $1 \le r \le n-1$, either $i_{r+1} \neq i_r$ or $i_{r+1} = i_r$ but $\varepsilon_{r+1} \neq -\varepsilon_r$; the empty word is also called reduced. Suppose that w is not reduced, and choose r such that $i_{r+1} = i_r$ and $\varepsilon_{r+1} = -\varepsilon_r$. Let w' be the word obtained from w by deleting the adjacent pair of letters $x_{i_r}^{\varepsilon_r}$ and $x_{i_{r+1}}^{\varepsilon_{r+1}}$. We say that w' comes from w by an *elementary reduction*. If w'' is obtained from w by a sequence of elementary reductions we say that w'' comes from w by *reduction*. It is usually convenient to allow w'' to be w in this definition (corresponding to the empty sequence of elementary reductions); readers will be left to decide for themselves on each occasion whether this case is permitted or not.

Examples Let w be the word $zxx^{-1}zy^{-1}y$. Then both $zzy^{-1}y$ and $zxx^{-1}z$ come from w by elementary reductions, and zz comes from w by reduction.

Let w be $zxx^{-1}xy$. Then zxy comes from w by elementary reduction. Two elementary reductions may be applied to w, both giving the same result; the first deletes the pair of letters xx^{-1}, while the second deletes the pair $x^{-1}x$.

Write, for the moment, $w \approx w'$ iff either w is identical to w' or there is a sequence of words $w_1, \dots, w_k$ for some k such that w_1 is w and w_k is w' and, for each $j < k$, one of w_{j+1} and w_j comes from the other by elementary reduction. Plainly $\approx$ is an equivalence relation. We denote the set of equivalence classes by $F(X)$. Whenever we have an equivalence relation on a set, the equivalence class of the element w will be denoted by $[w]$. This notation will be used without further comment (but occasionally the notation will have a different meaning; for instance, $[a,b]$ is used for a closed interval of real numbers).

It is easy to see that if u, v, w, and w' are words such that $w \approx w'$ then $uwv \approx uw'v$. It follows, since $\approx$ is an equivalence relation, that if $u \approx u'$ and $w \approx w'$ then $uw \approx u'w'$ (as both are equivalent to uw'). This enables us to define a multiplication on $F(X)$, by requirng $[u][w]$ to be $[uw]$. This multiplication is

plainly associative, and has an identity (the class of the empty sequence) which we denote by 1 (just as the empty sequence itself is also denoted by 1). When w is the word $x_{i_1}^{\varepsilon_1} \dots x_{i_n}^{\varepsilon_n}$ we let w^{-1} denote the word $x_{i_n}^{-\varepsilon_n} \dots x_{i_1}^{-\varepsilon_1}$. Then $ww^{-1} \approx 1$, as we see easily. Hence $F(X)$ is a group. Define a function $i: X \to F(X)$ by $xi = [x]$. Plainly $F(X)$ is generated by Xi.

Theorem 3 *$(F(X), i)$ is free on X.*

Proof Let G be a group and $f: X \to G$ a function. Then f extends to a function from $M(X \cup \bar{X})$ to G sending the word $x_{i_1}^{\varepsilon_1} \dots x_{i_n}^{\varepsilon_n}$ to the element $(x_{i_1}f)^{\varepsilon_1} \dots (x_{i_n}f)^{\varepsilon_n}$. It is immediate that if w' comes from w by an elementary reduction then w and w' have the same image. It follows that if $w \approx w''$ then w and w'' have the same image. Hence we may define a function φ from $F(X)$ to G by requiring $[w]\varphi$ to be the image of w. It is obvious that φ is a homomorphism and that $xf = xi\varphi$. Further, since Xi generates $F(X)$, there can be at most one homomorphism from $F(X)$ to G with specified values on Xi.//

Clearly, because elementary reduction decreases length, when we start with any word and apply elementary reductions one after another in an arbitrary way until no more can be applied, we will ultimately reach a reduced word. In particular, every equivalence class contains at least one reduced word.

Theorem 4 (Normal form theorem for free groups) *There is exactly one reduced word in each equivalence class.*

Remarks As a special case of this theorem, we find that there is exactly one reduced word which can be obtained by reduction from a given word.

Because this theorem is so important, we give several proofs. Similar proofs apply in many situations.

We have seen that every equivalence class contains at least one reduced word. The difficult property is that each class contains at most one reduced word, and we now give proofs of this fact.

Proofs (I) Canonical reduction.

Define a function $\lambda: M(X \cup \bar{X}) \to M(X \cup \bar{X})$ inductively by

$1\lambda = 1$, $x^{\varepsilon}\lambda = x^{\varepsilon}$, $(ux^{\varepsilon})\lambda = (u\lambda)x^{\varepsilon}$ if $u\lambda$ does not end in $x^{-\varepsilon}$, and $(ux^{\varepsilon})\lambda = v$ if $u\lambda$ is $vx^{-\varepsilon}$.

It is easy to check, by induction on the number of letters in w, that $w\lambda$ is reduced for any word w, that $w\lambda$ is w if w is reduced, and that $w\lambda$ comes from w by reduction. Also, because $u\lambda$ is reduced, $(ux^{\varepsilon}x^{-\varepsilon})\lambda = u\lambda$ for any u. It then follows, by induction on the length of v, that $(ux^{\varepsilon}x^{-\varepsilon}v)\lambda = (uv)\lambda$ for all v; that is, $w\lambda = w'\lambda$ if w' comes from w by elementary reduction. From this we see that $w\lambda = w''\lambda$ if $w \approx w''$. In particular, if w and w'' are reduced words with $w \approx w''$ then $w = w\lambda = w''\lambda = w''$, as required.

(II)The Diamond Lemma.

We first show that if both w' and w'' come from w by elementary reduction then either w' is the same as w'' or there is a word w^* which comes from both w' and w'' by elementary reductions.

If w' and w'' come from w by different elementary reductions then there are two possibilities (interchanging w' and w'' if necessary), both of which were illustrated in the examples earlier. The first is that w is $ux^{\varepsilon}x^{-\varepsilon}x^{\varepsilon}v$ for some (possibly empty) words u and v, and that w' is obtained by deleting $x^{\varepsilon}x^{-\varepsilon}$ and w'' is obtained by deleting $x^{-\varepsilon}x^{\varepsilon}$. In this case w'' is the same as w'. The second possibility is that w is $ux^{\varepsilon}x^{-\varepsilon}ty^{\delta}y^{-\delta}v$ for some (possibly empty) words t, u, and v, and that w' is $uty^{\delta}y^{-\delta}v$ and w'' is $ux^{\varepsilon}x^{-\varepsilon}tv$. In this case the word utv is obtained by elementary reduction from both w' and w''.

Now take two equivalent words w and w' and consider a sequence $w_1, \ldots, w_n$ such that w_1 is w, w_n is w', and, for each i, one of w_{i+1} and w_i comes by elementary reduction from the other. Suppose that there is some r such that w_{i+1} comes by elementary reduction from w_i for all $i < r$ and w_i comes from w_{i+1} by elementary reduction for all $i \geq r$. Then w_r comes from both w and w' by reduction.

If there is no such r then there must be some k such that both w_{k-1} and w_{k+1} come from w_k by elementary reduction. In this case, the argument above tells us that we may obtain a new sequence which also shows the equivalence of w and w', either by deleting w_k and w_{k+1} or by replacing w_k by a word w^* such that both w_{k-1} and w_{k+1} reduce to w^*. Since the sum of the lengths of the members of the sequence decreases when we make this change, it follows by induction that there must be some word which comes from both w and w' by reduction.

In particular, if w and w' are equivalent reduced words then they must be the same.

A slightly more complicated version of this proof is given in the exercises, which has the advantage that it can be applied in more general situations.

(III) van der Waerden's method.

This method is one we shall use in several later theorems.

Let S be the set of all reduced words, and let G be the group of all permutations of S. We shall define a homomorphism $\varphi: F(X) \to G$ such that $[w]\varphi$ acting on the empty sequence ( ) give the sequence w whenever w is a reduced word. In particular, if w and w' are reduced words with $w \approx w'$ then $[w] = [w']$ and so w and w', being the results of acting on ( ) by $[w]\varphi$ and $[w']\varphi$, must be the same.

Since we have already shown that $F(X)$ is free on X, we can obtain φ by defining a function $f: X \to G$. We define xf as the permutation sending w to wx if w does not end in x^{-1} and sending w to u if w is ux^{-1}. It is easy to check that this is a permutation of S, whose inverse sends w to wx^{-1} if w does not end in x and sends w to v if w is vx. This holds because a reduced word cannot end in xx^{-1} or $x^{-1}x$.

If w is the reduced word $x_{i_1}^{\varepsilon_1} \dots x_{i_n}^{\varepsilon_n}$ then $[w]\varphi$ is the product of the permutations $(x_{i_r} f)^{\varepsilon_r}$ and, inductively, the reuslt of acting on ( ) by $[w]\varphi$ is just w itself, as needed.

(IV) For the final method, we show that if w is a reduced word distinct from 1 then there is a homomorphism φ from $F(X)$ to a finite group such that $[w]\varphi$ is not the identity. Since $[1]\varphi$ is the identity, we cannot have $w \approx 1$.

Suppose that this has been shown, and let u and v be reduced words such that $u \approx v$; then $uv^{-1} \approx 1$. The word uv^{-1} need not be reduced, but if u and v are different it is easy to see that by reduction from uv^{-1} we will obtain a reduced word w different from 1. Further, $w \approx uv^{-1} \approx 1$, which is prohibited by the previous paragraph.

So let w be the reduced word $x_{i_1}^{\varepsilon_1} \dots x_{i_n}^{\varepsilon_n}$. We will define a homomorphism φ into the group S_{n+1} of permutations of $\{1, \dots, n+1\}$ such that $[w]\varphi$ sends 1 into $n+1$, whence $[w]\varphi$ is not the identity. To do this we require a function $f: X \to S_{n+1}$ such that $(x_{i_r})^{\varepsilon_r}$ sends r into $r+1$ for all $r \le n$.

So we need xf to send r to $r+1$ if x_{i_r} is x and $\varepsilon_r = 1$ and to send $r+1$ to r if x_{i_r} is x and $\varepsilon_r = -1$. If these conditions define a one-one function from some subset of $\{1, \dots, n+1\}$ to another subset we can extend this function to a permutation of $\{1, \dots, n+1\}$, and then define xf to be this permutation.

Can this definition require us to send r to two different numbers? This could only occur if x_{i_r} were x with $\varepsilon_r = 1$ and $x_{i_{r-1}}$ were also x with $\varepsilon_{r-1} = -1$. This cannot happen since w is a reduced word. Also this function could

only map two different numbers to s if $x_{i_{s-1}}$ were x and $\varepsilon_{s-1} = 1$ and x_{i_s} were also x with $\varepsilon_s = -1$, which is again impossible as w is reduced.//

We can now obtain a new proof of Proposition 2(iii).

Corollary *$i: X \to F(X)$ is injective.*

Proof If x and y are distinct elements of X then they are distinct reduced words. Hence they lie in different equivalence classes.//

Exercise 1 Prove the claimed results about canonical reduction. In particular, show that $w\lambda$ is always reduced, that $w\lambda$ is w if w is reduced, and that $w\lambda$ comes from w by reduction. Show also that $(ux^{\varepsilon}x^{-\varepsilon})\lambda$ is the same as $u\lambda$, and, by induction on the number of letters in v, that $(ux^{\varepsilon}x^{-\varepsilon}v)\lambda$ is the same as $(uv)\lambda$.

Exercise 2 Let w be $ux^{\varepsilon}x^{-\varepsilon}v$. If ux^{ε} is reduced we say that uv comes from w by leftmost reduction. Show that $w\lambda$ comes from w by repeating leftmost reduction until we get a reduced word.

We may prove the Normal Form Theorem, starting from the first result in the Diamond Lemma approach, by methods which apply to very general reduction relations where the previously given induction on length cannot be used. This is done in the exercises below.

Exercise 3 Let u and v be obtained by reduction from w. Show that there is a word w^* which is obtained from both u and v by reduction. (An indication of the argument is given in Figure 1. Here the top diamond exists by the Diamond Lemma approach, the left-hand diamond exists inductively, and then the right-hand diamond also exists inductively.)

Exercise 4 Let $w \approx w'$, and let $w_1, \ldots, w_n$ be a sequence with w_1 being w and w_n being w' such that, for all i, one of w_{i+1} and w_i is an elementary reduction of the other. If w_{k-1} and w_{k+1} are elementary reductions of w_k we say k is a peak. Use Exercise 3 to show that if the sequence has a peak then there is another sequence joining w to w' with fewer peaks. (Figure 2 provides a sketch of the argument.) Deduce the Normal Form Theorem.

Exercise 5 It is possible to give a proof of the Normal Form Theorem by the Diamond Lemma approach without using induction on the number of peaks. Consider the relation between words in which u is related to v iff there is a word w such that both u and v reduce to w. Using Exercise 3, show that this is an equivalence relation. Deduce that this relation coincides with the relation $\approx$.

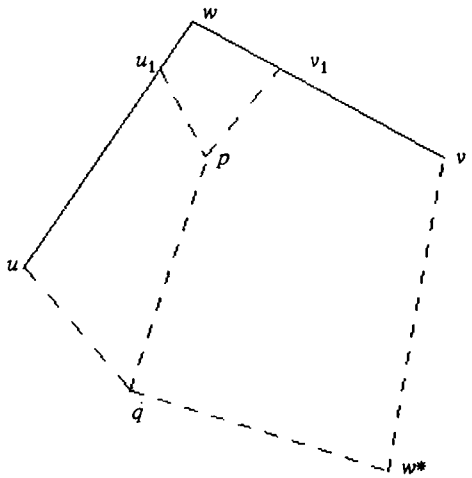

Figure 1

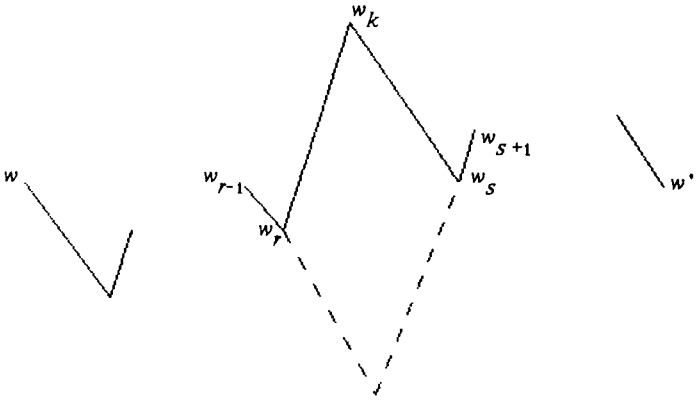

Figure 2

Exercise 6 Show that the function from the set of reduced words to itself which sends w to wx if w does not end in x^{-1} and which sends vx^{-1} to v is a permutation of the set of reduced words.

We usually regard X as a subset of $F(X)$ with i being inclusion; consequently we shall usually omit mention of i in future. We frequently identify elements of $F(X)$ with the corresponding reduced words. At times we need to regard words as elements of $M(X \cup \bar{X})$ and at other times we want to

regard them as giving elements of $F(X)$. We write $w \equiv w'$ when w and w' are the same word, while we write $w = w'$ if they define the same element of $F(X)$ (that is, if they are equivalent words; we no longer use the notation $\approx$).

It is easy to check, without using the normal form theorem, that if u and v are reduced words then there is only one sequence of elementary reductions which can be applied to uv to obtain a reduced word. Let u be $x_{i_1}^{\varepsilon_1} \dots x_{i_n}^{\varepsilon_n}$ and let v be $x_{j_1}^{\delta_1} \dots x_{j_m}^{\delta_m}$. Take $s \geq 0$ as large as possible such that for all $r \leq s$ we have $i_{n+1-r} = j_r$ and $\varepsilon_{n+1-r} = -\varepsilon_r$. Then the pairs $x_{i_{n+1-r}}^{\varepsilon_{n+1-r}} x_{j_r}^{\varepsilon_r}$ for $r = 1, \dots, s$ are successively deleted in reducing uv. The reduced word we obtain is therefore $x_{i_1}^{\varepsilon_1} \dots x_{i_{n-s}}^{\varepsilon_{n-s}} x_{j_{s+1}}^{\delta_{s+1}} \dots x_{j_m}^{\delta_m}$. It follows that it would be possible to define $F(X)$ to be the set of all reduced words with the product of u and v being given by the above formula. Some authors use this approach, but it is not really satisfactory. There is a practical problem, in that the proof of associativity is surprisingly messy with this definition (see Exercise 7) There is a more important theoretical objection to this approach, as it confuses the question of the existence of a free group and the question of the nature of normal forms for the elements. We frequently find in algebraic situations that it is easy to prove the existence of a free object but extremely difficult (sometimes impossible) to find normal forms for the elements.

Exercise 7 Suppose that we define $F(X)$ to be the set of reduced words, with the product of u and v being the unique reduced word obtained from uv by reduction. (We have seen that this word is unique without using the Normal Form Theorem.) Show that this product is associative.

A totally different approach may be used to show the existence of free groups. This approach works in other algebraic situations. It is of a categorical nature, amounting to the fact that we are looking for an adjoint to the forgetful functor. Unfortunately the proof runs into some set-theoretical difficulties. To make the proof easier, I first give the argument in the (incorrect!) form ignoring these difficulties, and then show how to resolve them.

Consider all pairs (f, G_f) where G_f is a group and f is a function from X to G_f. Take the cartesian product over all f of the groups G_f, which we call K. The set-theoretic problem arises here, since in formal set theory this product cannot be constructed because there are too many functions f; this problem will be ignored for the moment. There is a function $i: X \to K$ such that the component of xi in the factor G_f is xf. Let F be the subgroup of K generated by Xi. Let f be any function from X to a group. Then this group is