

# Generalized symmetry

J.B. Wilker

Let  $X$  be a set and  $G$  a group which acts on  $X$  and is generated by two elements  $a$  and  $b$ . Motivated by a geometric problem of L. Fejes Tóth, we define a subset  $S \subset X$  to have  $[a, b]$ -symmetry if its images under  $a$  and  $b$  satisfy  $S^a \cap S^b = S$ . The problem of finding all sets with  $[a, b]$ -symmetry when an arbitrary 2-generator group  $G$  acts on an arbitrary space  $X$  is shown to be equivalent to the same problem in the special case when the 2-generator free group acts on itself by right translation. This action is modelled in the hyperbolic plane in a way that helps to reveal the  $[a, b]$ -symmetric subsets of the free group.

## 1. The concept of generalized symmetry

Let  $X$  be a set and  $G$  a group acting on  $X$ . Let  $f, g, h$  be elements of  $G$  and  $S$  a subset of  $X$ . We say that  $S$  has generalized symmetry of the first type if  $S^f \cap S^g = S^h$  and  $S$  has generalized symmetry of the second type if  $S^f \cup S^g = S^h$ . Here  $S^f$  denotes the image of  $S$  under the permutation of  $X$  which corresponds to  $f$ .

The term generalized symmetry is appropriate because if  $g = h = e$ , the identity, the first equation becomes

$$S^f \cap S = S \quad \text{or} \quad S^f \supset S$$

and the second equation becomes

$$S^f \cup S = S \quad \text{or} \quad S^f \subset S.$$

---

Received 31 January 1978.

Taken together, these imply that  $S^f = S$  or that  $f$  is an ordinary symmetry of  $S$ .

Our two types of generalized symmetry are closely related. By taking complements through the equations which define them and making use of the fact that  $(S^f)' = (S')^f$  and so on, we see that  $S^f \cap S^g = S^h$  implies  $(S')^f \cup (S')^g = (S')^h$  and *vice versa*. Since a set has one type of generalized symmetry precisely when its complement has the other, we restrict attention to the first type of generalized symmetry. One might object that this precludes consideration of sets with both types of generalized symmetry. In point of fact, nothing new is lost by this restriction because a doubly symmetric set  $S$  must satisfy  $S^f = S^g = S^h$  and this indicates that  $f, g$ , and  $h$  are ordinary symmetries of  $S$  or ordinary symmetries followed by a common mapping.

An arbitrary element  $k \in G$  can be applied to both sides of the equation  $S^f \cap S^g = S^h$  to yield the equivalent equation  $S^{fk} \cap S^{gk} = S^{hk}$ . Our condition for generalized symmetry is therefore homogeneous and can be brought to canonical form by taking  $k$  to be  $f^{-1}, g^{-1}$ , or  $h^{-1}$ . We adopt the canonical form with  $k = h^{-1}$  so that the defining equation reads  $S^a \cap S^b = S$  with  $a = fh^{-1}$  and  $b = gh^{-1}$  playing equivalent roles. When this equation holds we say that  $S$  has  $[a, b]$ -symmetry and we write  $S \in [a, b]$ .

There are two different problems associated with generalized symmetry. In the first we are given  $a$  and  $b \in G$  and we seek the most general  $S \subset X$  such that  $S \in [a, b]$ . In the second, we are given  $S \subset X$  and we seek the most general  $a$  and  $b \in G$  such that  $S \in [a, b]$ . This paper treats only the first problem. Because of this, there is no point in taking  $G$  to be any larger than the group generated by  $a$  and  $b$ . In what follows, we assume  $G = \langle a, b \rangle$ .

## 2. Motivation

The motivation for these considerations is a question put to me by L. Fejes Tóth. In his question,  $X$  is euclidean  $n$ -space,  $g = T$  is a translation,  $h = K$  is a dilatation, and we are asked to find the most

general set  $S$  which satisfies  $S \cap S^T = S^K$ . In order to mention a few examples, we allow a taste of the second problem to flavour our discussion.

If  $S$  is a simplex and  $T$  is any translation such that  $S \cap S^T$  has non-void interior, then there is a  $K$  such that  $S \cap S^T = S^K$ . In fact this property can be used to characterize the simplex (see [2], [5], and references cited there). If  $S$  is a hypercube and  $T$  is translation part of the way along a body diagonal then there is a  $K$  such that  $S \cap S^T = S^K$ . The first indication that much more pathological sets can also possess this kind of symmetry is the observation that if  $S$  is the usual Cantor set on the real line,  $T$  is translation by  $2/3$  in either direction and  $K$  is the appropriate dilatation with scale factor  $1/3$ , then  $S \cap S^T = S^K$ .

In [5] there is a fairly complete discussion of the problem of Fejes Tóth. In that paper we refer to solutions of  $S \cap S^T = S^K$  by writing  $S \in (T, K)$  rather than  $S \in [K^{-1}, TK^{-1}]$ . This is quite natural considering the statement of the problem. However the present notation seems more appropriate because of its greater symmetry and its connection with various solutions to the problem. Many geometrically appealing sets satisfying  $S \cap S^T = S^K$  are double cones with their vertices at the fixed points of the dilatations  $K^{-1}$  and  $TK^{-1}$ . Notice that the simplex and the hypercube both admit this description.

Further connections with [5] will be mentioned as we progress. This paper is, in effect, an elaboration of the general considerations which proved significant in [5].

### 3. The $[a, b]$ -hull

If the sets  $S_i$  ( $i \in I$ ) all satisfy  $S_i^a \cap S_i^b = S_i$  then so does their intersection,  $S = \cap\{S_i : i \in I\}$ . Given any subset  $R \subset X$  we can define

$$\bar{R} = \cap\{S : S \supset R \text{ and } S \in [a, b]\} .$$

Since  $X \in [a, b]$ ,  $\bar{R}$  is well defined. Evidently  $\bar{R}$  is the smallest set containing  $R$  and belonging to  $[a, b]$ . We refer to  $\bar{R}$  as the  $[a, b]$ -hull of  $R$ .

In [5] we obtained the following description of  $\bar{R}$ . Let  $R_0 = R$  and for  $n \geq 0$  let

$$R_{n+1} = R_n \cup R_n^{a^{-1}} \cup R_n^{b^{-1}} \cup \left( R_n^a \cap R_n^b \right).$$

Then  $\bar{R} = \bigcup_{n=0}^{\infty} R_n$ . One corollary of our first theorem is an improved description of  $\bar{R}$  prompted by a remark of M.F. Newman.

Recall that  $G = \langle a, b \rangle$  is the group generated by  $a$  and  $b$ . Let  $G^- \subset G$  be the semigroup with identity generated by  $a^{-1}$  and  $b^{-1}$ . For each  $n \geq 1$ , let  $G^{-n} \subset G^-$  be the set of at most  $2^n$  elements which can be expressed as words of length  $n$  in  $a^{-1}$  and  $b^{-1}$ . If  $X_0 \subset X$  and  $G_0 \subset G$ , let

$$X_0^{G_0} = \left\{ x^g : x \in X_0 \text{ and } g \in G_0 \right\}.$$

**THEOREM 1.** *Let  $S$  be a subset of  $X$ . Then  $S \in [a, b]$  if and only if*

- (i) *for all  $x \in X$ ,  $x \in S$  implies  $x^{G^-} \subset S$ ,*
- (ii) *for all  $x \in X$  and  $n \geq 1$ ,  $x^{G^{-n}} \subset S$  implies  $x \in S$ .*

**Proof.** The condition  $S^a \cap S^b = S$  is obviously equivalent to

- (i)'  $S^a \cap S^b \supset S$ , and
- (ii)'  $S^a \cap S^b \subset S$ .

Now (i)' is equivalent to  $S^a \supset S$  and  $S^b \supset S$ , and therefore to  $S \supset S^{a^{-1}}$  and  $S \supset S^{b^{-1}}$ . Since  $G^-$  is a semigroup with identity this last pair of conditions is equivalent to  $S = S^{G^-}$  and this in turn is

equivalent to condition (i) of the theorem.

Condition (ii)' is equivalent to say, for all  $y, z \in S$ ,  $y^a = z^b = x$  implies  $x \in S$  and this is equivalent to saying, for all  $x \in X$ ,  $x^{a^{-1}} \in S$  and  $x^{b^{-1}} \in S$  implies  $x \in S$  or, alternatively, for all  $x \in X$ ,  $x^{G^{-1}} \subset S$  implies  $x \in S$ . This is the case  $n = 1$  of condition (ii) of the theorem. An easy induction shows that condition (ii) is self-improving and therefore that the general condition is equivalent to the special case with  $n = 1$ .

**COROLLARY 1.** *If  $R$  is a subset of  $X$  then  $\bar{R}$ , the  $[a, b]$ -hull of  $R$ , is given by*

$$\bar{R} = R^{G^-} \cup \{x \in X : \exists n \geq 1 \text{ with } x^{G^{-n}} \subset R^{G^-}\}.$$

*Proof.* Any set with  $[a, b]$ -symmetry which contains  $R$  must contain  $R_1 = R^{G^-}$  by Theorem 1 (i) and then must contain

$$R_2 = \{x \in X : \exists n \geq 1 \text{ with } x^{G^{-n}} \subset R^{G^-}\}$$

by Theorem 1 (ii). It follows that  $R_1 \cup R_2 \subset \bar{R}$ . To prove that  $R_1 \cup R_2 = \bar{R}$  we must verify that it has  $[a, b]$ -symmetry and we do this by checking conditions (i) and (ii) of Theorem 1.

If  $x \in R_1$ ,  $x^{G^-} \subset R_1^{G^-} = R_1$  and if  $x \in R_2$  with  $x^{G^{-n}} \subset R_1$ , the infinite tail of  $x^{G^-}$  lies in  $R_1$  and the finitely many elements in  $\bigcup_{m=1}^{n-1} x^{G^{-m}}$  lie in  $R_2$ . It follows that  $R_1 \cup R_2$  satisfies condition (i).

If  $x^{G^{-n}} \subset R_1 \cup R_2$  then the  $2^n$  points  $x_i \in x^{G^{-n}}$  must satisfy

$x_i^{G^{-m_i}} \subset R_1$  for some  $m_i \geq 1$ . If  $m = \max\{m_1, m_2, \dots, m_{2^n}\}$  then

$x^{G^{-(n+m)}} \subset R_1$  and  $x \in R_2$ . It follows that  $R_1 \cup R_2$  satisfies condition

(ii) as well.

**COROLLARY 2.** *Suppose that for all  $x \in X$ , there exists  $m = m(x) \geq 1$  such that  $x \in G^{-m}$ . Then if  $R$  is a subset of  $X$ ,  $\bar{R} = R^{G^-}$ .*

*Proof.* If  $x \notin R_1 = R^{G^-}$ , then for arbitrarily large values of  $k$ ,  $x^{G^{-km}} \notin R_1$ , because  $x \in x^{G^{-km}}$ . The proof of Corollary 1 indicates that if a point  $x$  is added to  $\bar{R}$  via  $R_2$ , then for all sufficiently large values of  $n$ ,  $x^{G^{-n}} \in R_1$ . Our opening remark therefore shows that if  $x \notin R_1$ , then  $x \notin R_2$ , and hence  $x \notin \bar{R}$ .

**COROLLARY 3.** *Suppose that for all  $x \in X$  and  $m \geq 1$ ,  $x \notin G^{-m}$ . Then if  $R$  is a subset of  $X$ ,*

$$\bar{R} = R^{G^-} \cup \{x \in X : |x^{G^-} - R^{G^-}| < \infty\}.$$

*Proof.* Under the conditions of the corollary the expression

$$x^{G^-} = \bigcup_{n=0}^{\infty} x^{G^{-n}}$$

is a disjoint union. It follows that  $|x^{G^-} - R^{G^-}| < \infty$  if and only if  $x^{G^{-n}} \in R^{G^-}$  for sufficiently large values of  $n$ .

#### 4. An example

Let  $X$  be the euclidean plane and let  $a$  and  $b$  be the translations  $(x, y) \rightarrow (x, y) + (1, 0)$  and  $(x, y) \rightarrow (x, y) + (0, 1)$  respectively. The orbit of  $(0, 0)$  under the group  $G = \langle a, b \rangle$  is the integer lattice  $Z^2$ , and we seek the most general set  $S \subset Z^2$  such that  $S^a \cap S^b = S$ .

The semigroup  $G^-$  consists of all integer translations  $(x, y) \rightarrow (x, y) + (s, t)$  with  $s \leq 0$  and  $t \leq 0$ . The set  $G^{-n}$  reduces to the  $n + 1$  translations of this type with  $s + t = -n$ . It follows

that  $(0, 0)^{G^-}$  is equal to the third quadrant of  $Z^2$  and if  $(x_0, y_0) \in Z^2$ ,  $(x_0, y_0)^{G^-}$  is equal to the third quadrant translated by  $(x_0, y_0)$ . These sets possess  $[a, b]$ -symmetry and constitute the minimal non-void examples.

In order to find the most general set with  $[a, b]$ -symmetry we can use Theorem 1, Corollary 3. If  $\emptyset \neq R \subset Z^2$ ,  $R_1 = R^{G^-}$  is a union of translates of the third quadrant and a line of  $Z^2$  parallel to a coordinate axis either misses it, meets it in a negatively directed ray, or is contained in it. If  $x = c$  meets  $R_1$ , then so does  $x = c'$  for  $c' \leq c$ . Since  $R_1 \neq \emptyset$  it is clear that  $x = c$  meets  $R_1$  as  $c \rightarrow +\infty$ . We define

$$c_0 = \sup\{c \in Z : x = c \text{ meets } R_1\}$$

and, on the basis of similar considerations,

$$d_0 = \sup\{d \in Z : y = d \text{ meets } R_1\}.$$

If  $c_0$  and  $d_0$  are both equal to  $+\infty$ , then for any  $(x_0, y_0) \in Z^2 - R_1$ , the lines  $x = x_0$  and  $y = y_0$  meet  $R_1$  at  $(x_0, y_1)$  and  $(x_1, y_0)$ . The set  $(x_0, y_0)^{G^-} - R_1$  contains at most  $(x_0 - x_1)(y_0 - y_1)$  points,  $(x_0, y_0) \in R_2$ , and it follows that  $\bar{R} = Z^2$ .

If  $c_0$  and  $d_0$  are both finite, then  $(c_0, d_0) \in R_2$ , but points of the form  $(c_0 + 1, y)$  or  $(x, d_0 + 1)$  do not belong to  $R_2$ . It follows that  $\bar{R}$  is equal to the third quadrant translated by  $(c_0, d_0)$ .

If  $c_0$  is finite and  $d_0 = +\infty$ , points of the form  $(c_0, y)$  all belong to  $R_2$ , but those of the form  $(c_0 + 1, y)$  do not. It follows that  $\bar{R}$  is equal to the halfplane  $x \leq c_0$ . A similar consideration applies if  $c_0 = +\infty$  and  $d_0$  is finite.

An instructive modification of this example is to roll the integer lattice into a cylinder  $Z \times Z_m$  by identifying  $(x, y)$  and  $(x, y+m)$  for some  $m \geq 1$ . The transformation  $a$  continues to be aperiodic but  $b$  now satisfies  $b^m = e$ . Every point of the cylinder satisfies

$(x, y) \in (x, y)G^{-m}$  and Theorem 1, Corollary 2, applies. The  $[a, b]$ -symmetric sets are  $\emptyset$ , the half-infinite cylinders  $x \leq c_0$ , and the whole cylinder.

### 5. The orbit partition

If  $G$  is a group acting on a space  $X$ , the sets  $x^G$  ( $x \in X$ ) are called the orbits of  $G$ . Two orbits either coincide or are disjoint and so they give a partition of  $X$ ,  $X = \cup\{X_i : i \in I\}$ . In our case  $G = \langle a, b \rangle$  and the orbits  $X_i$  ( $i \in I$ ) all satisfy  $X_i^a = X_i^b = X_i$  and therefore possess  $[a, b]$ -symmetry. It follows that if  $S$  is an arbitrary set with  $[a, b]$ -symmetry then its orbital parts  $S_i = S \cap X_i$  all possess  $[a, b]$ -symmetry. The most general set  $S \in [a, b]$  can therefore be written in the form

$$S = \cup\{S_i : i \in I\},$$

where  $S_i \subset X_i$  and  $S_i \in [a, b]$ .

It is not in general true that if the sets  $S_j$  ( $j \in J$ ) all satisfy  $S_j^a \cap S_j^b = S_j$  then so does their union,  $S = \cup\{S_j : j \in J\}$ . However this is true if

$$\cup\{S_j^a \cap S_k^b : j, k \in J\} \subset \cup\{S_j^a \cap S_j^b : j \in J\},$$

and in particular it is true if  $S_j^a \cap S_k^b = \emptyset$  whenever  $j \neq k$ . One way to achieve this is to let  $S_j \subset X_j$  where  $X = \cup\{X_j : j \in J\}$  is the  $G$ -orbit partition of  $X$ . For then it follows that  $S_j^a \cap S_k^b \subset X_j^a \cap X_k^b = X_j \cap X_k$ , and hence  $S_j^a \cap S_k^b = \emptyset$  whenever  $j \neq k$ . This shows that any set of the



form given at the end of the last paragraph must possess  $[a, b]$ -symmetry.

**THEOREM 2.** *Let  $X = \cup\{X_i : i \in I\}$  be the  $\langle a, b \rangle$ -orbit partition of  $X$ . Then a subset  $S \subset X$  belongs to  $[a, b]$  if and only if*

$$S = \cup\{S_i : i \in I\}$$

where  $S_i \subset X_i$  and  $S_i \in [a, b]$ .

**COROLLARY.** *If  $X$  has cardinality  $c$  of the continuum, it possesses  $2^c$  subsets with  $[a, b]$ -symmetry.*

**Proof.** The group  $G = \langle a, b \rangle$  has at most  $\aleph_0$  elements. It follows that each orbit has at most  $\aleph_0$  points and so there must be  $c$  orbits. On each orbit we have at least the 2-fold choice  $S_i = X_i$  or  $S_i = \emptyset$  and this shows that there are at least  $2^c$  subsets of  $X$  with  $[a, b]$ -symmetry. On the other hand  $X$  has only  $2^c$  subsets!

Theorem 2 was mentioned in [5] for the sake of its corollary. It is important in our present treatment because it shows that we can solve the problem of  $[a, b]$ -symmetry orbit by orbit. This is very convenient because we have a useful model of the way in which a group acts on an orbit.

### 6. Transitive action

Let  $G$  be a group and  $H$  a subgroup. Let  $X = G/H = \{Hg : g \in G\}$  be the set of right cosets of  $H$  in  $G$ . Then  $G$  acts on  $X$  by right translation: an element  $r \in G$  maps  $Hg \in X$  to  $Hgr \in X$ . There is only one orbit under this action because  $Hg_1$  is mapped to  $Hg_2$  by

$g_1^{-1}g_2$ . The subgroup  $H \subset G$  is the stabilizer of the point  $H \in X$ .

It turns out that this is the only way a group  $G$  can act transitively on a set  $X$ . To see this, choose a base point  $x_0 \in X$  and let  $H$  be the subgroup of  $G$  which fixes  $x_0$ . If  $x \in X$  and  $g \in G$  satisfy  $x_0^g = x$  then the coset  $\bar{x}_0(x) = Hg$  can be assigned as a coordinate to name the point  $x$ . This coordinate coset can be described

as the set of all elements of  $G$  which map  $x_0$  to  $x$ . It follows that the action of  $G$  on  $X$  is equivalent to right translation on these right cosets. For if  $r \in G$  maps  $x$  to  $y$  then  $gr$  maps  $x_0$  to  $y$ ,  $Hgr$  is the coordinate for  $y$ , and the mapping of coordinates carries  $Hg$  to  $Hgr$ . This observation goes back to Burnside ([1], Chapter IX).

The correspondence between points  $x \in X$  and right cosets  $\bar{x}_0(x) \subset G$  extends to a correspondence between subsets  $S \subset X$  and unions of right cosets  $T \subset G$  where  $T = \bar{x}_0(S) = \cup\{\bar{x}_0(x) : x \in S\}$ . This extended correspondence commutes with the action of  $G$ : if  $S$  corresponds to  $T$  then  $S^r$  corresponds to  $Tr$ . The mapping  $\bar{x}_0$  from subsets of  $X$  to coordinatizing subsets of  $G$  has an inverse which we will call  $\overline{\bar{x}_0}$ . A moment's thought shows that  $\overline{\bar{x}_0}$  is given by the elegant expression

$$\overline{\bar{x}_0}(T) = x_0^T.$$

Let us change  $G$ -coordinates on  $X$  from the base point  $x_0$  to the base point  $y_0$ . If  $y_0^k = x_0$  and the stabilizer of  $x_0$  is  $\bar{x}_0(x_0) = H$ , then the stabilizer of  $y_0$  is the conjugate subgroup  $\bar{y}_0(y_0) = kHk^{-1}$ . If  $x \in X$  satisfies  $x = x_0^g$ , then it also satisfies  $x = y_0^{kg}$ , and its  $y_0$ -based coordinate is

$$\bar{y}_0(x) = (kHk^{-1})kg = kHg = k\bar{x}_0(x).$$

It follows that the coordinate of a subset  $S \subset X$  changes from  $\bar{x}_0(S)$  to  $\bar{y}_0(S) = k\bar{x}_0(S)$ .

In our case  $G = \langle a, b \rangle$  might not act transitively on  $X$  but we can always obtain transitive action by passing to the individual orbits. For simplicity, we temporarily assume that our  $G$  does act transitively on  $X$  and we carry on with the notation of the preceding paragraphs. We have remarked that the coordinate function  $\bar{x}_0$  commutes with the action of  $G$ .

It follows that a subset  $S \subset X$  satisfies  $S^a \cap S^b = S$  if and only if the

subset  $\bar{x}_0(S) = T \subset G$  satisfies  $Ta \cap Tb = T$ .

If a subset  $T \subset G$  has  $[a, b]$ -symmetry, then so does  $kT$  for any  $k \in G$ . This follows immediately if we multiply the equation  $Ta \cap Tb = T$  on the left by  $k$ . If  $T = \bar{x}_0(S)$ , one point of view is that the various sets  $kT$  are the different names for  $S$  corresponding to different base points in  $X$ . However if  $k$  lies in the normalizer of  $H$ , there is an alternative point of view. Under this condition, if  $T = \cup\{Hg_j : j \in J\}$ , then  $kT = \cup\{kHg_j : j \in J\} = \cup\{Hkg_j : j \in J\}$ . Since  $kT$  is a union of right cosets of  $H$ ,  $\bar{x}_0(kT)$  is a well defined  $[a, b]$ -symmetric subset of  $X$ , and it will be different from  $S$  if  $kT$  is different from  $T$ .

**THEOREM 3.** *Let  $G = \langle a, b \rangle$  act transitively on  $X$ , let  $x_0$  be a point of  $X$ , and let  $H$  be the subgroup of  $G$  which fixes  $x_0$ . There is a one-to-one correspondence between the subsets  $S \subset X$  which satisfy  $S^a \cap S^b = S$  and the subsets  $T \subset G$  which satisfy  $Ta \cap Tb = T$  and are equal to a union of right cosets of  $H$ . Under this correspondence the set  $T$  is given by*

$$T = \bar{x}_0(S) = \left\{ g \in G : x_0^g \in S \right\}$$

and the set  $S$  is given by

$$S = \bar{x}_0(T) = x_0^T.$$

The example of Section 4 provides a simple illustration of Theorem 3. When  $G = \langle a, b \rangle = Z^2$  acts on itself the subsets  $T \subset G$  which satisfy  $Ta \cap Tb = T$  are  $\emptyset$ , translates of the third quadrant, halfplanes  $x \leq c_0$  and  $y \leq d_0$  and  $G$  itself. When  $G$  acts on the cylinder  $Z \times Z_m$ , the subgroup  $H$  is equal to  $\langle b^m \rangle$  at every base point. The  $[a, b]$ -symmetric subsets of  $G$  which can be expressed as a union of right cosets of  $H$  are  $\emptyset$ , the halfplanes  $x \leq c_0$ , and  $G$  itself. These give rise to the  $[a, b]$ -symmetric subsets of  $Z \times Z_m$ .

## 7. The free group

When a group  $G$  acts on a set  $X$  there is a homomorphism of  $G$  into  $\text{perm } X$ . Related to Theorem 2 is the fact that  $X$  admits a partition into  $G$ -orbits,  $X = \cup\{X_i : i \in I\}$ . Related to Theorem 3 is the fact that once we choose a base point  $x_i \in X_i$ , then  $X_i$  can be modelled as a right coset space  $X_i = G/H_i$ . A transformation of  $G$  fixes every point in the orbit  $X_i$  if it lies in  $K_i$ , the intersection of the stabilizers  $gH_i g^{-1}$  ( $g \in G$ ). A transformation of  $G$  fixes every point in  $X$  if it lies in  $K = \cap\{K_i : i \in I\}$ . Thus  $K$  is the kernel of the homomorphism of  $G$  into  $\text{perm } X$ .

In our case when  $G = \langle a, b \rangle$  acts on  $X$ , it is important to notice that the  $[a, b]$ -symmetric subsets of  $X$  are determined not by  $G$  but by its homomorphic image in  $\text{perm } X$ . Suppose  $F = \langle \alpha, \beta \rangle$  has  $G$  as a homomorphic image and this homomorphism carries  $a \rightarrow \alpha$  and  $b \rightarrow \beta$ . The group  $F$  acts on  $X$  through its homomorphism onto  $G$  and the image of  $F$  in  $\text{perm } X$  is the same as that of  $G$ . It follows that  $[\alpha, \beta]$ -symmetric subsets of  $X$  are the same as  $[a, b]$ -symmetric subsets of  $X$ . The utility in this is that Theorem 3 can be applied with  $F$  in the role of  $G$ . If  $F$  is chosen to be the free group on two generators then  $F$  maps homomorphically onto every 2-generator group  $G$ . Theorem 4 is an alternative to Theorem 1, Corollary 1, in providing a universal answer to the question of which sets possess  $[a, b]$ -symmetry.

**THEOREM 4.** *Let  $F = \langle \alpha, \beta \rangle$  be the free group on two generators. The problem of finding the  $[a, b]$ -symmetric subsets of an arbitrary set  $X$  is equivalent to the problem of finding the subsets  $T \subset F$  which satisfy  $T\alpha \cap T\beta = T$  and  $T = \cup\{Hf_j : j \in J\}$  for arbitrary subgroups  $H \subset F$ .*

## 8. Geometry of the free group

It may appear that our abstract approach to the question of generalized symmetry has taken us rather far from its geometric origin. This is not the case. In Section 4 we saw that the free abelian group on two generators acting on itself by right translation is equivalent to the group generated by two perpendicular translations in the euclidean plane

acting on any orbit. In this section we shall see that the free group on two generators acting on itself by right translation is equivalent to the group generated by two sufficiently long, perpendicular translations in the hyperbolic plane acting on any orbit. This observation was prompted by a remark of B.H. Neumann.

He pointed out that the 4-valent graph of the free group on two generators can be represented very elegantly within the square  $|x| + |y| \leq 1$ . The edges of this graph are all parallel to the coordinate axes but they are not all of the same length. The origin is a vertex and its nearest neighbours are the four points  $(\pm\frac{1}{2}, 0)$  and  $(0, \pm\frac{1}{2})$ . Each of these vertices has for its nearest neighbours the origin and three new points a distance  $\frac{1}{2}$  away. This pattern continues so that each vertex at graph theoretic distance  $n - 1$  from the origin is joined to three vertices at graph theoretic distance  $n$  from the origin by edges of length  $1/2^n$ .

Figure 1 shows this graph with its edges "coloured" and directed to make its automorphism group equal to the free group on generators  $a$  and  $b$ . This group is sharply transitive on the vertices of the graph, so once one vertex is labelled  $e$ , the remaining vertices  $v$  can be labelled by the automorphism which moves  $e$  to  $v$ . When this is done, the action of any automorphism can be determined by right translating the vertex labels.

Figure 2 shows a topologically equivalent graph drawn in the unit circle  $|z| < 1$ . Adjacent edges meet at right angles and are represented by arcs of circles perpendicular to  $|z| = 1$  so that they correspond to straight line segments in the Poincaré model of the hyperbolic plane. The polygons joining  $e$  to  $a$  to  $ba$  to  $a^{-1}ba$ , and so on, and  $e$  to  $b$  to  $ab$  to  $b^{-1}ab$ , and so on, have edges of equal length if and only if their vertices lie on a cycle. These two polygons are guaranteed not to interfere with one another if this cycle is a horocycle or hypercycle rather than a circle. The minimum edge length arises in the case of the horocycle and is equal to  $\log(1+\sqrt{2})$ , which is twice the distance whose angle of parallelism is  $\pi/4$ . In this case the vertex labelled  $a$  is located at  $z = \sqrt{2}/2$  and the one labelled  $b$  is located at  $z = (\sqrt{2}/2)i$ . The resulting graph has an automorphism group consisting of hyperbolic

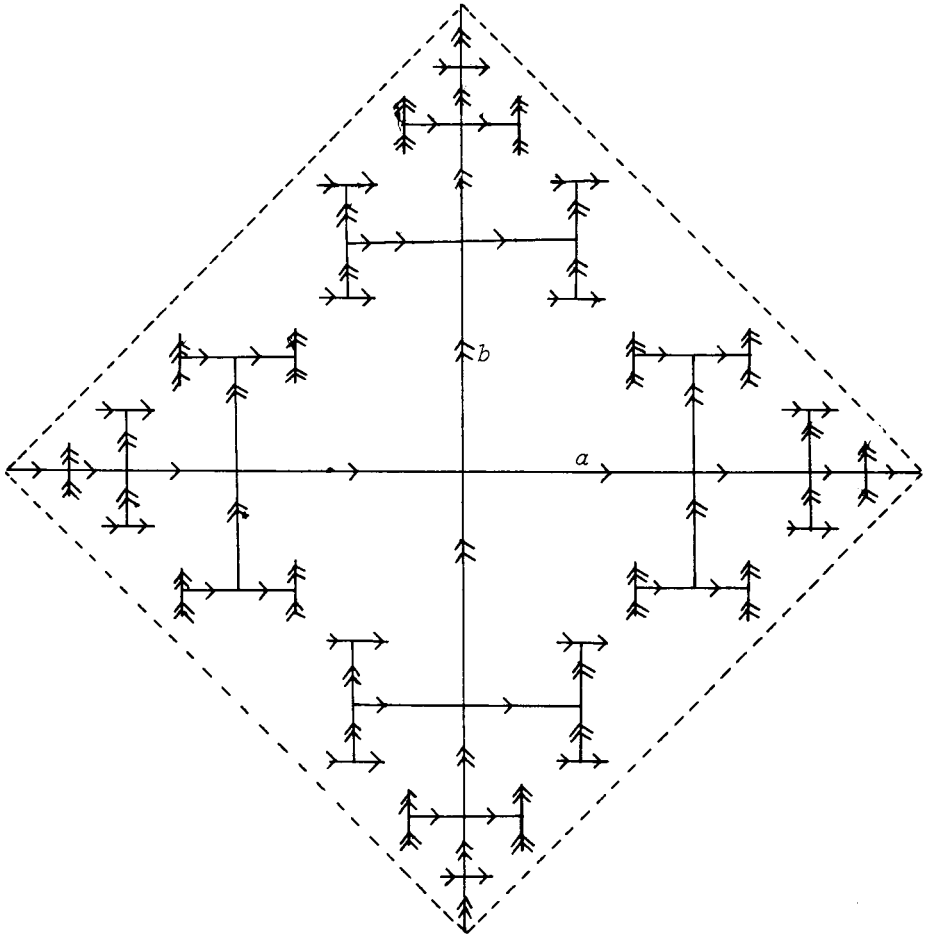


FIGURE 1

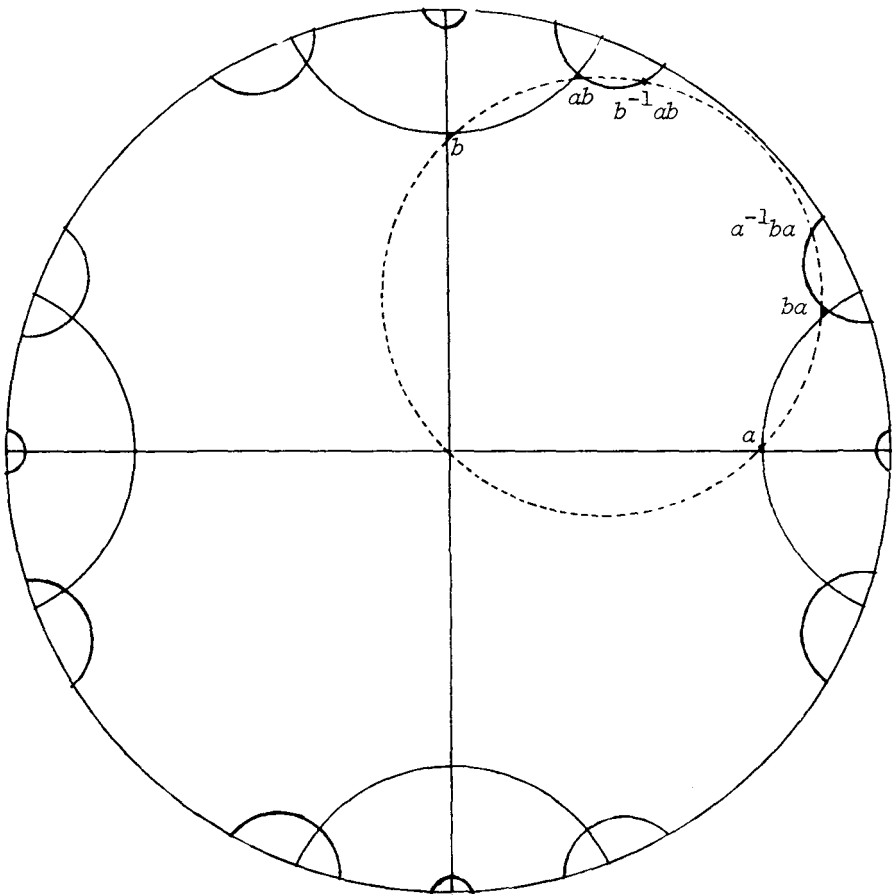


FIGURE 2

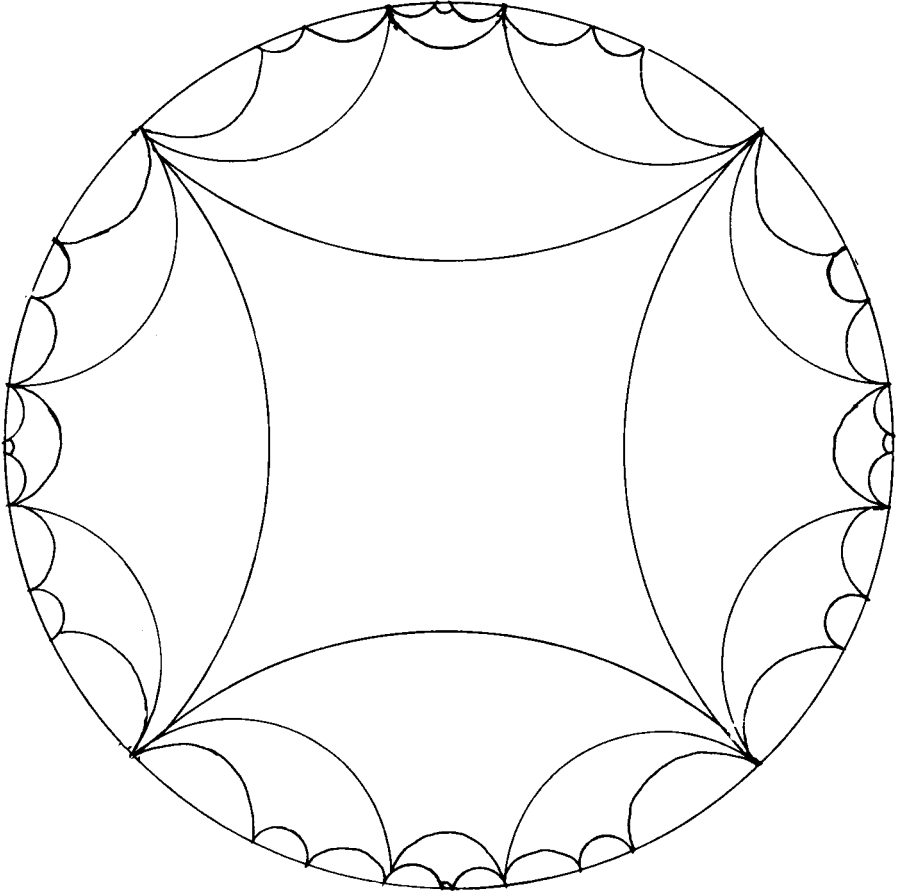


FIGURE 3



isometries. By colouring and directing its edges as in Figure 1 we reduce this automorphism group to the free group on two generators. These generators are the perpendicular translations  $a$  and  $b$  given by

$$z \rightarrow \frac{2z + \sqrt{2}}{\sqrt{2}z + 2} \quad \text{and} \quad z \rightarrow \frac{2z + \sqrt{2}i}{-\sqrt{2}iz + 2} .$$

It is clear that the group generated by  $a$  and  $b$  acts without fixed points on the orbit of  $z = 0$  because these points comprise the vertices of our graph. The group transports the open disk of hyperbolic radius  $\frac{1}{2} \log(1 + \sqrt{2})$  centred at  $z = 0$  to congruent disks centred at the other vertices of our graph. Since these disks do not overlap, the group acts without fixed points on the orbit of any point in their union. We are led to consider a fundamental region for the group generated by  $a$  and  $b$ . This can be taken equal to the asymptotic quadrangle shown in Figure 3. Since this quadrangle tiles the hyperbolic plane without overlap, the group generated by  $a$  and  $b$  acts without fixed points on every orbit. This means that the action on every orbit is equivalent to the group acting on itself by right translation. We remark that our Figure 3 is essentially the same as Magnus' Figure 19a ([3], p. 178), and therefore its history goes back to 1882.

### 9. Subsets of the free group

A subset  $T$  of the free group  $F$  generated by  $a$  and  $b$  has  $[a, b]$ -symmetry if  $Ta \cap Tb = T$ . Theorem 1, Corollary 3, applies in this context to say that if  $R \subset F$ , its  $[a, b]$ -hull is given by

$$\bar{R} = RF^{-1} \cup \{x \in F : |xF^{-1}RF^{-1}| < \infty\} .$$

The graph of the free group has just been discussed, and if we modify the labelling given a moment ago, we can obtain an effective graphical procedure for computing  $\bar{R}$ .

Our original labelling of Figure 1 is adapted to right translations. Thus the vertex labelled  $w$  has neighbours  $aw$  to the east,  $a^{-1}w$  to the west,  $bw$  to the north and  $b^{-1}w$  to the south. If the labels are all reversed without changing the convention about how  $a$  and  $b$  act on the graph, we obtain a new labelling which is adapted to left translation. In this labelling the vertex labelled  $u$  has neighbours  $ua$  to the east,

$ua^{-1}$  to the west,  $ub$  to the north, and  $ub^{-1}$  to the south. Regardless of the labelling of vertices it is natural to distinguish the directed edges of the graph by calling them  $a, a^{-1}, b$ , or  $b^{-1}$ , according to whether they proceed west to east, east to west, south to north, or north to south. The left action labelling of vertices has the advantage that if the symbols for the edges of the polygon from  $e$  to  $u$  are taken in their natural order they spell out  $u$ .

Now consider the left action labelling of the graph of  $F$  together with the formula for computing  $\bar{R}$  in  $F$ . The set  $F^-$  consists of all points which can be reached from  $e$  by steps directed south or west. We say that these points lie south-west of  $e$  and take special note of the fact that they comprise a vanishingly thin fraction of the third quadrant points. In this descriptive language the set  $RF^-$  consists of all points south-west of a point of  $R$  and  $xF^-$  consists of all points south-west of  $x$ . It is not hard to form an intuitive picture of those  $x$  for which  $|xF^- - RF^-| < \infty$ .

By using this graphical technique, we see that if  $R = \{a^{-1}, b^{-1}\}$ , then the set  $R_1 = RF^- \dagger [a, b]$ . The set  $R_2$  adds the required point  $e$  to give  $\bar{R} = F^-$ . In spite of this simple example, there is a useful class of sets  $R$  with the property that  $\bar{R} = R_1 = RF^-$ .

**THEOREM 5.** *Let  $F = \langle a, b \rangle$  be the free group generated by  $a$  and  $b$  and let  $F^+ \subset F$  be the semigroup with identity generated by  $a$  and  $b$ . Then if  $R \subset F^+$ , the  $[a, b]$ -hull of  $R$  is given by  $\bar{R} = RF^-$ .*

*Proof.* The negative semigroup  $F^-$  can be written in the form

$$F^- = \{e\} \cup F^-a^{-1} \cup F^-b^{-1}.$$

If  $S = RF^-$ , then  $S$  satisfies

$$\begin{aligned} Sa &= RF^-a \\ &= R[\{e\} \cup F^-a^{-1} \cup F^-b^{-1}]a \\ &= Ra \cup RF^- \cup RF^-b^{-1}a \\ &= S \cup [R \cup RF^-b^{-1}]a, \end{aligned}$$

and similarly

$$Sb = S \cup [R \cup RF^{-1}a^{-1}]b .$$

Since  $R \subset F^+$  ,  $[R \cup RF^{-1}b^{-1}]a$  consists of reduced words which end in  $a$  , and  $[R \cup RF^{-1}b^{-1}]b$  consists of reduced words which end in  $b$  . It follows that  $Sa \cap Sb = S$  and since  $R \subset S \subset \bar{R}$  ,  $\bar{R} = S$  .

**COROLLARY.** *Let  $F = \langle a, b \rangle$  be the free group generated by  $a$  and  $b$  . Then  $F$  contains  $c$  distinct subsets  $T$  which satisfy  $Ta \cap Tb = T$  .*

**Proof.** There are  $c$  real numbers  $t$  satisfying  $0 < t < 1$  . Each such real number can be associated with a distinct binary "decimal" and hence, by changing 0's to  $a$ 's and 1's to  $b$ 's , with a distinct infinite sequence of  $a$ 's and  $b$ 's . Let  $R_t \subset F^+$  be the subset whose elements are initial segments of the series corresponding to  $t$  . Let  $T_t = R_t F^-$  . According to Theorem 5, the sets  $T_t$  all possess  $[a, b]$ -symmetry. Since  $T_t \cap F^+ = R_t$  , it follows that  $T_{t_1} \neq T_{t_2}$  if  $t_1 \neq t_2$  . This gives us  $c$  different  $[a, b]$ -symmetric subsets of  $F$  , and since  $F$  has a total of  $2^{\aleph_0} = c$  subsets, the result follows.

This corollary stands in contrast to the result of Section 4 where we showed that the free abelian group with generators  $a$  and  $b$  has only  $\aleph_0$   $[a, b]$ -symmetric subsets.

### 10. Generalized symmetry in classical geometry

Let  $X$  be euclidean  $n$ -space and  $a$  and  $b$  two affinities, or let  $X$  be hyperbolic, spherical, or elliptic  $n$ -space and  $a$  and  $b$  two isometries. Find the most general set  $S$  which satisfies  $S^a \cap S^b = S$  and describe the classes of these sets which are geometrically appealing. This is the full problem which I would like to solve in further work under the heading of generalized symmetry.

To solve the first part of this problem and find the most general set which satisfies  $S^a \cap S^b = S$  , let  $X = U\{X_i : i \in I\}$  be the  $\langle a, b \rangle$ -orbit partition of  $X$  and use Theorems 2, 3, and 4 together with the graphical method of finding  $T$ 's which satisfy  $Ta \cap Tb = T$  in the free group  $F$  .

The difficulty which this prescription conceals is the difficulty of determining the groups  $G = \langle a, b \rangle$  and expressing their orbits in the form  $X_i = F/H_i$ . The following is a first step towards doing this.

**THEOREM 6.** *Let  $G = \langle a, b \rangle$  be a 2-generator subgroup of the affinities of euclidean  $n$ -space or the isometries of hyperbolic, spherical or elliptic  $n$ -space. Then for almost every point  $x$  of the space,  $G$  acts on the orbit  $x^G$  as if by right translation on itself.*

**Proof.** The group  $G$  has at most countably many elements. Each of these except the identity fixes a set of points of measure 0. The union of these countably many fixed point sets is therefore a null set. It follows that almost every point  $x$  lies in the complementary set and is stabilized only by the identity. Since stabilizers along an orbit are conjugate subgroups, this means that almost every point  $x$  lies in an orbit  $x^G$  on which  $G$  acts without fixed points. This proves the result.

Theorem 6 allows us to neglect the atypical orbits if we are content to find our  $[a, b]$ -symmetric subsets up to measure zero. However it leaves unanswered the question of which abstract groups  $G = F/H$  can arise and how. The next step in our continuing program is to collect information on this question.

The problem of determining the finitely generated subgroups of the classical geometric groups is a natural extension of the problem of determining their finite subgroups. This problem is of interest in a broader context than that of generalized symmetry. For example, the construction which yields the Hausdorff-Banach-Tarski Paradox ([4], Chapter XI) depends on the fact that the infinite group

$G = \langle a, b : a^2 = b^3 = e \rangle$  can arise in euclidean 3-space generated by a halfturn and a  $120^\circ$  rotation about intersecting axes.

## References

- [1] W. Burnside, *Theory of groups of finite order* (Cambridge University Press, Cambridge, 1897).

- [2] René Fourneau, "Nonclosed simplices and quasi-simplices", *Mathematika* **24** (1977), 71-85.
- [3] Wilhelm Magnus, *Noneuclidean tessellations and their groups* (Pure and Applied Mathematics, 61). Academic Press, New York and London, 1974).
- [4] Herbert Meschkowski, *Unsolved and unsolvable problems in geometry* (translated by Jane A.C. Burlak. Oliver and Boyd, Edinburgh, London; Frederick Ungar, New York; 1966).
- [5] J.B. Wilker, "Sets homothetic to their intersection with a translate", *Canad. Math. Bull.* **18** (1975), 739-748.

Department of Mathematics,  
Institute of Advanced Studies,  
Australian National University,  
Canberra, ACT.