A Lemma that is not Burnside's

Peter M. Neumann

The Queen's College, Oxford, England

(Received: 11 September 1978)

1. The lemma

The fact which is to be the subject of my comments in this note is very well known because, besides having many applications in, pure group theory, it lies at the foundation of that part of combinatorics which deals with counting problems and is associated with the name of Pólya (or Redfield, or both). It refers to the situation where we have a finite set $\Omega$ (of, for want of a better term, 'points') on which a group $G$ acts as a group of permutations (not necessarily faithfully); we deem points $\alpha, \beta$ to be related when there exists $g$ in $G$ such that $\alpha^g = \beta$; this is an equivalence relation whose equivalence classes are known as the 'orbits' of $G$ in $\Omega$. The lemma states that the number of orbits is the average of the numbers of points fixed by elements of $G$; in a notation which should be self-explanatory,

$$\frac{1}{|G|} \sum_{g \in G} |\text{fix}_\Omega(g)| = |\text{orb}_\Omega(G)|.$$  \hspace{1cm} (1)

The notation and terminology which I have used here is that of pure permutation group theory but it can easily be interpreted in other contexts. As a familiar example consider the problem of calculating the number of ways in which the six faces of a cube may be coloured using, say, at most three colours, red, white and blue. Here we are not interested in distinguishing between colourings which are 'essentially the same' in the sense that after a suitable rotation of the one coloured cube it looks just the same as the other. In this case we take $\Omega$ to be the set of all possible colourings of a fixed cube, of which there are $3^6$, and we take $G$ to be the group of all rotations of the cube (of which there are just 24: we are considering only proper rotations, not also reflections) acting on our set $\Omega$ in the obvious way. Two colourings are 'essentially the same' precisely if the corresponding members of $\Omega$ are related by a rotation, and so 'orbits' as defined above correspond exactly to the 'essentially different ways' of colouring our cube. To use Equation (1) we need to understand the members of $G$ well enough to know how many colourings each one leaves fixed. The following list gives the information in such a form
that you can, I hope, work out how it was obtained:

| Description of rotations                           | Number of such rotations | Number of colourings fixed | Total contribution to $\sum |\text{fix}_G(g)|$ |
|----------------------------------------------------|--------------------------|----------------------------|----------------------------------|
| Identity rotation:                                  | 1                        | $3^6$                       | 729                              |
| Axis through pair of opposite vertices of cube;    |                          |                             |                                  |
| Angle $2\pi/3$ or $4\pi/3$:                         | 8                        | $3^2$                       | 72                               |
| Axis through midpoints of pair of opposite edges;  |                          |                             |                                  |
| Angle $\pi$:                                        | 6                        | $3^3$                       | 162                              |
| Axis through centres of pair of opposite faces;    |                          |                             |                                  |
| Angle $\pi$:                                        | 3                        | $3^4$                       | 243                              |
| Axis through centres of pair of opposite faces;    |                          |                             |                                  |
| Angle $\pi/2$ or $3\pi/2$:                         | 6                        | $3^3$                       | 162                              |
| Significant totals:                                 | 24                       |                             | 1368                             |

Division of 1368 by 24 to get the average shown on the left side of Equation (1) gives 57 as the number of varieties of cubes coloured with red, white and blue. Of course, to get the number of colourings that really use all three colours (if that had been the original problem) we would have to subtract 3 for the monochrome cubes and the number of ways that use just two of the colours. This number can be obtained in just the same way: I find that 27 of the colourings do not use all the colours and so there are just 30 essentially different ways of colouring the cube red, white and blue.

I hope there are readers of this journal to whom the lemma is not already thoroughly well known because I wish to rehearse the argument before commenting on it. Briefly, one may see the proof as taking two steps. The first and more substantial is to deal with the case where $G$ acts ‘transitively’ on $\Omega$: that is, for every $\alpha, \beta$ in $\Omega$ there exists $g$ in $G$ such that $\alpha^g = \beta$. This is the case, in other words, where there is just one orbit. Then we let

$$S := \{(\omega, g) \mid \omega \in \Omega, g \in G, \omega^g = \omega\},$$

and we count the members of $S$ in two different ways. One way gives that $|S| = \sum_{g \in G} |\text{fix}_G(g)|$, the other that $|S| = \sum_{\omega \in \Omega} |G_\omega|$, where $G_\omega := \{g \in G \mid \omega^g = \omega\}$. Now we find that for any $\alpha \in \Omega$ the set $G_\alpha$ is a subgroup of $G$ (known as the ‘stabiliser’ of $\alpha$);
furthermore, if $\beta \in \Omega$ then $\{ g \in G \mid \alpha^g = \beta \}$ is a coset $G_{\alpha}x$ of $G_{\alpha}$ in $G$; consequently, since $G$ is supposed to act transitively, the set $\Omega$ can be put into one–one correspondence with the set of all cosets $G_{\alpha}x$; thus $|\Omega|$ is the number of these cosets, that is, the index $|G : G_{a}|$; but of course, each coset contains the same number $|G_{a}|$ of elements of $G$ and so we see that $|\Omega| \cdot |G_{a}| = |G|$. This tells us that all the subgroups $G_{\alpha}$ have the same size (one can easily prove more: that they are in fact all conjugate in $G$), and so, for any $\alpha \in \Omega$, we have

$$|\mathcal{S}| = \sum_{\alpha \in \Omega} |G_{\alpha}| = |\Omega| \cdot |G_{a}| = |G|.$$  

Consequently, when $G$ is transitive we have that

$$|G| = |\mathcal{S}| = \sum_{g \in G} |\text{fix}_G(g)|.$$  

The second step of the proof is now very easy: we let $\Omega_1, \ldots, \Omega_t$ be the orbits of $G$ in $\Omega$, observe that $\text{fix}_G(g)$ is the disjoint union of the sets $\text{fix}_{\Omega_i}(g)$, and conclude that

$$\sum_{g \in G} |\text{fix}_G(g)| = \sum_{i=1}^{t} \sum_{g \in G} |\text{fix}_{\Omega_i}(g)| = \sum_{i=1}^{t} |G_{i}| = t|G|,$$

that is, that

$$\frac{1}{|G|} \sum_{g \in G} |\text{fix}_G(g)| = t = |\text{orb}_G(G)|,$$

which is Equation (1).

2. Some variations

There are many agreeable variations on the theme we have just played. As there is little space in this article I have restricted my choice to just a few of those that are pertinent to the historical remarks that I wish to make.

For the first variant of Equation (1) we suppose given a group $G$ and actions of it on sets $\Omega_1$ and $\Omega_2$. The set $\Omega_1 \times \Omega_2$ of ordered pairs carries a natural $G$-action defined by the equation $(\omega_1, \omega_2)^g = (\omega_1^g, \omega_2^g)$. It is very easy to see that

$$\text{fix}_{\Omega_1 \times \Omega_2}(g) = \text{fix}_{\Omega_1}(g) \times \text{fix}_{\Omega_2}(g),$$

and so we can deduce from (1) the apparently much more general equation

$$\frac{1}{|G|} \sum_{g \in G} \varphi_1(g)\varphi_2(g) = |\text{orb}_{\Omega_1 \times \Omega_2}(G)|$$

in which, for brevity we have put $\varphi_i(g) := |\text{fix}_{\Omega_i}(g)|$. Equations (1) and (2) are, of course, equivalent: we have derived (2) from (1); and if in (2) we take $\Omega_1 := \Omega$ and $\Omega_2$ to be a singleton set $\{\gamma\}$ on which $G$ acts in the only possible way ($\gamma^g = \gamma$ for all $g$ in $G$) then we retrieve Equation (1).

Consider next the situation where we are given a group $S$ and two of its subgroups
The direct product $H \times K$ (as abstract group: this may not be a subgroup of $S$) then acts on $S$ by the rule $x^{(h,k)} := h^{-1} x k$ for all $x \in S$ and all $(h,k) \in H \times K$. Orbits of $H \times K$ in this action on $S$ are the double cosets $HxK$, and so if the number of double cosets is $|S: H, K|$ then, applying (1) with $S$ for $\Omega$ and $H \times K$ for $G$, we find that

\[
\left\{ \begin{align*}
\text{the number of solutions of the} \\
\text{equation } h^{-1} x k = x \text{ with} \\
h \in H, k \in K, x \in S \text{ is} \\
|H| \cdot |K| \cdot |S: H, K|.
\end{align*} \right. \tag{3}
\]

Although (3) has been obtained from a very special case of (1), and although it appears to be rather different in that it refers to the internal structure of some group $S$, we can again retrieve (1) from (3). To do that take $S$ to be the symmetric group consisting of all permutations of $\Omega$, and $H$ to be the stabiliser of a point $x$ in it. Since the symmetric group is certainly transitive on $\Omega$ we find, using the argument in the proof on page 134, that there is a one–one correspondence between $\Omega$ and the set of cosets of $H$ in $S$; moreover, if $\beta$ corresponds to $Hx$ (that is, $\beta = x^\sigma$) and if $s \in S$, then $\beta^s$ corresponds to the coset $Hxs$; it follows that the $G$-orbit containing $\beta$ corresponds to $\{Hxg | g \in G\}$, that is, to the set of cosets lying in the one double coset $HxG$ in $S$. In this situation therefore there is a one–one correspondence between the set of $G$-orbits in $\Omega$ and the set of double cosets $HxG$, and so

\[|S: H, G| = |\text{orb}_\Omega(G)|.\]

Now for a given element $g$ of $G$ the number of pairs $h, x$ with $h \in H, x \in S$ and $h^{-1} x g = x$ is $|\text{fix}_\Omega(g)| \cdot (n - 1)!$ because $x g x^{-1} \in H$ if and only if the permutation $x$ is such that $\sigma^x \in \text{fix}_\Omega(g)$. Therefore (3) tells us that

\[ (n - 1)! \sum_{g \in G} |\text{fix}_\Omega(g)| = |H| \cdot |G| \cdot |S: H, G|, \]

from which (1) follows immediately since $|H| = (n - 1)!$

### 3. Characters

Associated with the action of $G$ on $\Omega$ there is a matrix representation and its character. We associate with an element $g$ of $G$ the permutation matrix $P(g)$ whose rows and columns are indexed by $\Omega$, and in which the $(\alpha, \beta)$-entry is 1 if $\sigma^g = \beta$ and 0 if $\sigma^g \neq \beta$. The map $g \mapsto P(g)$ is then a matrix representation. By definition its character $\varphi$ is given by $\varphi(g) := \text{trace}(P(g))$. But the $(\alpha, \omega)$-entry of $P(g)$ is 0 unless $\sigma^g = \alpha$, in which case it is 1, and so

\[ \varphi(g) = |\text{fix}_\Omega(g)|. \]

We can now re-interpret Equation (1): if $\chi_0$ denotes the principal character of $G$ (that is, $\chi_0(g) = 1$ for all $g \in G$) then

\[ \frac{1}{|G|} \sum_{g \in G} \chi_0(g) \varphi(g) \]
is the usual inner product $\langle \chi_0, \varphi \rangle$ of characters of $G$, and (1) becomes

$$\langle \chi_0, \varphi \rangle = |\text{orb}_\Omega(G)|.$$  

(4)

This equation can also be proved directly. Let $K\Omega$ denote the vector space of all formal $K$-linear combinations of elements of $\Omega$ (here $K$ may as well be the field of complex numbers, although other fields will also do). Our group $G$ acts in an obvious way on $K\Omega$ permuting the basis elements and, as this action is $K$-linear, $K\Omega$ becomes a $KG$-module. The element $\sum_{i=1}^{n} a_i \omega_i$ of $K\Omega$ is fixed by every member of $G$ if and only if the coefficients $a_i$ are constant over each orbit of $G$ in $\Omega$. Thus, for the space of invariants of $G$ in $K\Omega$, we have

$$\dim \text{fix}_{K\Omega}(G) = |\text{orb}_\Omega(G)|,$$

(5)

and it is one of the basic facts of representation theory that the dimension on the left side of this equation is the inner product of $\chi_0$ with the character of the matrix representation afforded by $K\Omega$. Since this character is $\varphi$ we have Equation (4); this therefore has given us a character-theoretic interpretation, and actually another proof, of Equation (1).

This may seem rather a roundabout route to (1), but in fact (5) and equivalent forms of it, such as the apparently far more general equation

$$\dim \text{hom}_{K\Omega}(K\Omega_1, K\Omega_2) = |\text{orb}_{\Omega_1 \times \Omega_2}(G)|$$

(6)

(which bears the same relation to (5) as (2) does to (1)), allow one to use what one knows about the representation theory of finite groups over an arbitrary field to study permutation groups.

4. Commentary

My excuse for discussing at this length some of the ramifications of so well known a fact is that I wished to set the scene for some remarks about its origins. Nowadays the result is frequently, but quite incorrectly, referred to as ‘Burnside's Lemma’ or ‘Burnside's Theorem’. Of course, an incorrect attribution does harm to no-one because those who are not interested in the history of ideas will not notice, and those who do prefer to get such matters right can easily find out for themselves. Nevertheless, this particular misnomer appears to be of sufficiently recent origin that one may hope to correct it. Besides, there are one or two connections with other matters of some historical interest.

I have remarked that the mistake appears to be of quite recent origin, and perhaps it is worth a short digression to comment on this. A cursory search through the literature has not brought to light any reference earlier than 1960 in which Burnside’s name is mentioned. For example, Redfield (1927) quotes no source at all, Pólya (1937) refers (on p. 167) to Theorem 102 of the textbook by Speiser (1927), and in (1959) we find de Bruijn writing simply

We shall use a well-known device from the theory of permutation groups, used by Pólya in his proof of the fundamental theorem.
The first source I have found that attributes the result to Burnside was published in 1961: Golomb ([1961], p. 406) writes *tut court*

(This result is due to Burnside, and was rediscovered by Pólya, Gleason, et al.)

In (1964), p. 150, de Bruijn writes with appropriate caution

An essential part of Pólya's theory consists of a simple lemma, presumably published for the first time by Burnside ([3], Sec. 145, Theorem VII) (his reference [3] is my Burnside (1911)) but in his next paper (1963) (see (1963), p. 161 and (1964), p. 182 for evidence that the latter was written before the former) he speaks of 'Burnside's lemma'. From that time on, the name seems to have come into common use (see, for example, Liu (1968), de Bruijn (1971), Harary (1970)).

The lemma was certainly known long before Burnside's book was written. It appears quite explicitly in a paper by Frobenius ([1887], p. 318 in his collected works, Vol. 2) and even earlier in some of Cauchy's work (1845). Cauchy assumes that his group is transitive, but, as we saw in the first of the proofs that I gave above, it is a very small step from that to the general case. Indeed, Cauchy assumes that his group is *l* -fold transitive, which means that it is transitive in the natural action on \( \Omega^l \), the set of ordered sequences \( (\omega_1, \omega_2, \ldots, \omega_l) \) of \( l \) distinct elements of \( \Omega \). He uses \( M \) to denote the order \( |G| \) of \( G \), and \( H_l \) to denote the number of members of \( G \) which move precisely \( s \) points of \( \Omega \), so leave \( n - s \) points of \( \Omega \) unaltered, where \( n := |\Omega| \). With this notation he has the equations

\[
\begin{align*}
M &= H_n + H_{n-1} + H_{n-2} + \cdots + H_2 + 1, \\
M &= H_{n-1} + 2H_{n-2} + \cdots + (n-2)H_2 + n, \\
M &= 1.2H_{n-2} + \cdots + (n-3)(n-2)H_2 + (n-1)n, \\
M &= 1.2.3\cdots lH_{n-1} + \cdots + (n-l+1)\cdots (n-1)n.
\end{align*}
\]

(15)

Since \( s(s-1)\cdots (s-k+1) \) is the number of points of \( \Omega^l \) fixed by an element of \( G \) that leaves precisely \( s \) points of \( \Omega \) unaltered, these equations may be written in my notation as

\[
\sum_{g \in G} |\text{fix}_{\Omega^l(g)}(g)| = |G| \text{ for } 0 \leq k \leq 1.
\]

If \( G \) is \( l \) -fold transitive then it is certainly \( k \) -fold transitive whenever \( k \leq l \), and so these equations are instances of (1). But of course, taking \( l := 1 \) we retrieve Equation (1) in the case when \( G \) is transitive and, as I have already remarked, this is the heart of the matter.

Cauchy's proof is *au fond* very similar to the familiar proof that I rehearsed at the beginning of this note, but it is obscured by various complications. Instead of working simply with the order of \( G \) and the order and index of a stabiliser \( G_a \), Cauchy discusses the proportions of the members of each conjugacy class of \( \text{Sym}(\Omega) \) that lie in \( G \) and in the stabiliser \( G_a \). Furthermore, Cauchy's applications of his equations (15) were really rather ineffective, so that, even though these applications concerned one of the most fashionable problems of nineteenth-century mathematics (the search—which I intend to discuss in another essay elsewhere—for the possibilities for the index \( |S_n : G| \) of a subgroup \( G \) in the symmetric group \( S_n \)), his work seems to have been rather neglected.
by twentieth-century writers. For example, G. A. Miller, who presumably knew Cauchy's work well, since he wrote a considerable number of papers on the history of group theory and held that Cauchy was its founder, writes on p. 32 of the book Miller, Blichfeldt and Dickson (1916),

This interesting theorem was given explicitly for the first time by G. Frobenius, *Crelle*, vol. 101 (1887), p. 287.

This paper (1887) by Frobenius (who gives due credit to Cauchy, but who appears to have had his own ideas on the subject) is considerably more efficient than Cauchy's. He states Equation (1) as Theorem II of §4. Earlier in the paper he had proved what I have called (3) by combinatorial calculations inside the group $S$, and he deduces (1) from (3) by more or less the same argument as the one that I have given. Perhaps that is how he discovered the lemma, or perhaps he knew it from Cauchy's work and came to his own proof as a by-product of his study of double cosets. In any event, although he goes on to give some rather indecisive applications of the same kind as Cauchy's, Frobenius certainly seems to have understood the lemma and its importance better than Cauchy did, and he gives, as a second proof, the now familiar argument with which I began this note.

This work of Frobenius was followed very soon by a paper of Netto (1888), in which results about the average number of cycles of any length in a given group of permutations were published. Nowadays these are naturally subsumed into the general character theory of permutation groups. What one does is to apply Equation (1) or Equation (4) to suitable $G$-spaces derived from $\Omega$, such as the sets $\Omega^{(k)}$ of all ordered sequences of $k$ distinct elements of $\Omega$, or the sets $\Omega^{(k)}$ of all unordered such sequences. Under suitable conditions (such as, that $G$ is $k$-fold transitive on $\Omega$) this gives a considerable amount of information about the characters of $G$. Of course, these ideas post-date those papers of Frobenius and Netto: but not by long. Frobenius was developing character theory between 1895 and 1898, and although in his writing he is never very explicit about the relationship between enumeration theorems for permutation groups on the one hand, and characters on the other, he certainly seems to have been well aware of it. It is implicit in his 1897 paper and very close to the surface in the papers of 1900 and 1904, where he uses it to obtain the irreducible characters of the symmetric groups and other multiply transitive permutation groups. In Burnside (1900) and Frobenius (1901) character theory is used the other way round, to prove substantial theorems about permutation groups (the two-fold transitivity of insoluble transitive groups of prime degree, and the existence of regular normal subgroups in what are now called 'Frobenius' groups). Although Thomas Hawkins, in his study (1971), (1972), (1974) of the development of character theory, makes no mention of the relationship with permutation groups, I wonder if the experience that Frobenius acquired when he wrote his 1887 paper might not have influenced him later?

To return to Equation (1): my thesis is, that the result was so well understood early this century that it naturally found its way into the standard textbooks and monographs, mainly without attribution. Its appearance in Burnside's book is typical and is not to be taken as an indication that he was claiming it as his own discovery. If it is to be given a name it could be the 'Cauchy–Frobenius Lemma' but certainly not 'Burnside's Lemma'.
5. Acknowledgements

It is a pleasure to record my thanks to Mr Prabir Bhattacharya for help with recent bibliographical material, to Dr D. J. Gates and his colleagues at CSIRO who offered suggestions and editorial advice, and to Professor N. G. de Bruijn who, in a charming and helpful letter dated 28 August 1978 wrote, among other things,

Indeed, I think that I am to blame, having used the name “Burnside’s lemma” in several of my papers. You describe correctly how this all went. Pólya did not give a reference, Golomb mentioned the name Burnside, I looked it up in Burnside’s book and found it without a reference, so that was that. As a matter of fact I knew the lemma already (probably from Speiser’s book) when I first studied Pólya’s paper in about 1943, so my description as a “well-known device from the theory of permutation groups” was not bad at all. And as Burnside lived and worked at the time of the birth of group representation theory, this all fitted very well.

Since a few years I know the reference to Frobenius, who could certainly not have built up representation theory without it. But it is a great surprise that you have been able to trace it to Cauchy. I remember other cases where Cauchy had things that were rediscovered very much later; he was certainly at least half a century ahead of his time. The name “Cauchy–Frobenius lemma” seems to be well-chosen.

References


