THE STEINBERG REPRESENTATION

J. E. HUMPHREYS

To Robert Steinberg on his 65th birthday

Introduction. Group representations occupy a sort of middle ground between abstract groups and transformation groups, i.e., groups acting in concrete ways as permutations of sets, homeomorphisms of topological spaces, diffeomorphisms of manifolds, etc. The requirement that the elements of a group act as linear operators on a vector space limits somewhat the complexity of the action without sacrificing the depth or applicability of the resulting theory. As in other areas of mathematics, study of linear phenomena may illuminate more general phenomena.

The widespread use of group representations in mathematics (as well as in physics, chemistry, . . .) does not imply the existence of a single unified subject, however. Nor do practitioners always understand one another's language. Groups come in many flavors: finite, infinite-but-discrete, compact, locally compact, etc. Vector spaces may be finite or infinite dimensional; in the latter case there might be a Hilbert space structure and operators might be required to be unitary. The underlying scalar field may be complex, real, p-adic, finite, . . . . One can also make groups act on free modules over rings of arithmetic interest such as \( \mathbb{Z} \).

Even the study of finite group representations, which probably came first historically, has become somewhat fragmented. Traditionally one considers representations of finite groups by \( n \times n \) matrices with entries from \( \mathbb{C} \). These are the "ordinary" representations. But in the late 1930s Richard Brauer began to show the usefulness of "modular" representations (with matrix entries lying in a field of prime characteristic) as a tool in the ordinary theory and in the structure theory of finite groups. There is now an active modular industry, with a life of its own, benefiting from recent innovations such as quivers and almost split sequences in the representation theory of finite dimensional algebras (which include group algebras). Study of "integral" representations is equally active, motivated by number-theoretic considerations or by questions raised by topologists about integral group rings of fundamental groups.
The representation theory of compact groups imitates at first the finite case. It is still an essentially finite dimensional theory, with Haar integrals replacing finite sums. The main features of the ordinary representation theory of finite groups remain valid: complete reducibility, Schur's Lemma, orthogonality relations. But once one specializes to compact Lie groups, the theory (as pioneered by Elie Cartan and Hermann Weyl) takes off in new directions: highest weight of a representation, realization of irreducible representations with a fundamental dominant highest weight, Weyl character formula, invariant theory, unified treatment of special functions. The needs of modern physics push the subject beyond compact groups to locally compact groups, especially Lie groups: nilpotent, solvable, semisimple, reductive. But the representation theory becomes essentially infinite dimensional and far more intricate. Heavy analysis mixes with algebra in Harish-Chandra's far-reaching program, which is still at the forefront of contemporary mathematics. Similar ideas permeate the representation theory of $p$-adic Lie groups.

Many semisimple Lie groups are in fact linear algebraic groups (defined by polynomial equations). The finite dimensional "polynomial" representations of semisimple algebraic groups like $\text{SL}_n$ or reductive groups like $\text{GL}_n$ defined over fields of arbitrary characteristic all bear a strong family resemblance; but in prime characteristic, these usually fail to be completely reducible and take on many features of the infinite dimensional theory for Lie groups. Not content with finite dimensional groups, newer pioneers have found good reason to delve into representations of "Kac-Moody groups" associated with infinite dimensional Kac-Moody Lie algebras.

The study of group representations might be regarded as yet another example of hopelessly fragmented and overspecialized mathematics, were it not for the beauty and applicability of many of its ideas. But the sheer quantity of results and techniques is a deterrent to anyone who seeks a unified overview. Lie theory provides a glimmer of hope, since there is a surprising amount of unity here—at least, after the fact. For example, Harish-Chandra's philosophy of cusp forms in the infinite dimensional setting carries over largely intact to the ordinary characters of finite groups of "Lie type" in the Deligne-Lusztig theory. These are the groups such as $\text{PSL}_n(q)$ defined over finite fields which mimic semisimple or reductive Lie groups remarkably well and (together with alternating groups) yield most of the finite simple groups. There is also a close resemblance between the modular representation theory of these groups (in the defining characteristic) and the Cartan-Weyl theory of highest-weight representations. Just as the Kazhdan-Lusztig polynomials associated to Weyl groups have led to character formulas for infinite dimensional highest-weight modules, the analogous polynomials associated to affine Weyl groups are conjectured by Lusztig to yield character formulas for finite dimensional modular representations of groups of Lie type (and corresponding algebraic groups such as $\text{SL}_n$).

In spite of the underlying unity of much of this work, it is impossible to expose it adequately in a few pages. Instead, we focus on a single topic which conveys very well the flavor of much recent work—especially in prime characteristic.

Everyone who comes in contact with the representation theory of finite groups of Lie type quickly becomes aware of the Steinberg representation, a
distinguished representation of prime power degree [55]. Even though it is just one of many ordinary irreducible representations (over $\mathbb{C}$) or modular representations (over fields of the prime characteristic underlying the group), it seems to turn up with disproportionate frequency in all sorts of questions: see [48] for a recent instance. Moreover, it arises independently in a number of contexts, involving not just the finite groups but also the ambient algebraic groups and their Lie algebras. The purpose of the present exposition is to sort out these facets of the Steinberg representation for the nonspecialist.

Our viewpoint is to regard the Steinberg representation as a paradigm of a classical theorem due to Brauer and Nesbitt [12, Theorem 1].

**Brauer-Nesbitt Theorem.** Let $G$ be a group of order $p^a b$, where $p$ is prime and $(p, b) = 1$. An ordinary irreducible representation of degree divisible by $p^a$ remains irreducible after "reduction modulo $p"$, where it is also a "principal indecomposable" representation determining a "block" by itself. Moreover, the character of the representation vanishes at all elements of $G$ having order divisible by $p$.

To explain in detail what this theorem means and how the Steinberg representation illustrates it, we proceed in three steps. §1 describes Steinberg's construction of an ordinary representation of degree equal to the order of a Sylow $p$-subgroup in a finite group of Lie type. Character values are also discussed, together with alternate approaches and applications. §2 deals with the role of the Steinberg representation as an irreducible modular representation of both the finite group and its parent algebraic group. Lie algebra representations come into the picture as well. Finally, §3 relates all of this to the projective modules in characteristic $p$ (principal indecomposable modules) and resulting blocks (indecomposable two-sided ideals in the group algebra).

The finite group $G$ in question is realized concretely as a subgroup of an algebraic group defined over a finite field, the easiest example being $\text{SL}_n(q)$, $q = \text{power of a prime } p$. This particular group is of "universal" type, coming as it does from the simply connected algebraic group $\text{SL}_n(K)$, where $K$ is an algebraic closure of the prime field $\mathbb{F}_p$. For such a group there is a single Steinberg representation. Closely related groups of Lie type such as $\text{PGL}_n(q)$, $\text{PSL}_n(q)$, or $\text{GL}_n(q)$ can be discussed similarly, but they may give rise to extra characters of the kind found by Steinberg (e.g., after multiplying by a power of the character det).

As general background we can cite the survey article of Curtis [21] and the book of Carter [13], on representations over $\mathbb{C}$, as well as the survey [37] and the forthcoming book of Jantzen [38], on modular theory.

1. **St as an ordinary representation.**

1.1 *Steinberg's construction.* The existence of characters of prime power degree for finite groups of Lie type was perhaps first observed around 1900, in the work of Frobenius and Schur on characters of $\text{SL}_2(q)$. But not until the 1950s did such observations become systematic, in the early work of Steinberg on classical groups (cf. [54] and part I of [55]) and the independent work of Green on $\text{GL}_n(q)$. After the publication of Chevalley's famous Tôhoku paper on finite simple groups in 1955, Steinberg, Tits, and others rapidly completed
the list of simple groups of Lie type and devised axiomatic descriptions to facilitate their further study. In part II of [55], Steinberg constructed what is now called the Steinberg representation for the Chevalley groups. Later he adapted it to the twisted groups and studied the character values in more detail, cf. [56, 57].

To outline Steinberg’s construction, we need a little more notation. As in the Introduction, $G$ denotes a finite group of Lie type, which for convenience (and without significant loss of generality) we take to be of universal type. We also assume for ease of exposition that $G$ can be realized as the subgroup $\mathcal{G}(F_q)$ of some simple algebraic group $\mathcal{G}$ defined over $F_q$. But similar considerations apply to the groups of Ree, Suzuki, and Tits. Various subgroups of $\mathcal{G}$ then play a key role in representation theory, notably a Borel subgroup $B$ and a (split) maximal torus $T$ contained in $\mathcal{G}$, together with the Weyl group $W = J/\mathcal{G}$, where $J = N_G(T)$. $W$ is a finite group generated by reflections. The corresponding subgroups of $G$ are denoted $B$, $T$, $N$. These are the main ingredients in the axiomatic notion of $(B, N)$-pair or Tits system (cf. [13, 21]).

Note that in much of the literature, the “torus” $T$ is instead called a “Cartan subgroup” $H$.

In the case $\mathcal{G} = SL_n(K)$, we might take $T$ to be the subgroup of diagonal matrices, while $B$ might be the upper triangular matrices; then $W$ is the symmetric group on $n$ letters, with representatives inside $\mathcal{G}$ consisting of permutation matrices. In general, $B = T U$ (semidirect), where $U$ consists of unipotent matrices. Over $F_q$, unipotent just means $p$-singular: having order a power of $p$. The finite group $U$ turns out to be a $Sylow$ $p$-subgroup of $G$. One other bit of notation: Like the symmetric group, any Weyl group has a distinguished “alternating” character $\varepsilon$ with values 1 and $-1$.

Steinberg’s basic idea is to construct a (left) ideal of the group algebra of $G$ which affords the desired representation. At first he can work over an arbitrary field $F$. An explicit generator of the ideal is written concisely as:

$$e = \sum_{w \in W} \varepsilon(w) n_w \overline{T} \overline{U},$$

where $\overline{T}$ and $\overline{U}$ denote the sums over these groups inside $FG$ and where $n_w$ denotes some fixed representative of $w \in W$ inside $N$. The resulting left ideal is shown to have a vector space basis consisting of the elements $u \cdot e$ ($u \in U$), so the dimension is the order of $U$, namely $q^m$ (where $m = \dim \mathcal{G}$, the number of “positive roots”). Moreover, the matrices representing $G$ with respect to this basis have integer entries 0, 1, $-1$ (at most $|W|$ nonzero entries per row), and $U$ itself acts just as in its left regular representation. Provided the characteristic of $F$ does not divide $[G : B]$, e.g., $F = \mathbb{C}$ or $F = K$, the representation thus constructed is absolutely irreducible. (On the other hand, the representation fails to be irreducible if the characteristic divides $[G : B]$; cf. [58] for some discussion of this case.)

We shall denote the character of the Steinberg representation (and sometimes the representation itself) by $St$.

1.2 Character values. What does the construction outlined above reveal about the character values of $St$? First of all, they lie in $\mathbb{Z}$. Moreover, $St$
vanishes at \( p \)-singular elements other than 1, by comparison with the regular representation of \( U \). Of course, the Brauer-Nesbitt Theorem guarantees more generally that \( \text{St} \) vanishes except at \( p \)-regular elements: elements of order not divisible by \( p \). But the precise values of \( \text{St} \) at \( p \)-regular elements remain to be studied.

In [56] Steinberg computed these values by an indirect method using the representation theory of \( \mathcal{G} \) (cf. (2.4) below). If \( s \in G \) is \( p \)-regular, then the absolute value of \( \text{St}(s) \) is the power of \( p \) dividing the order of the centralizer \( C_G(s) \). This fits nicely with the degree formula \( \text{St}(1) = q^m \). However, the sign of \( \text{St}(s) \) is determined here only implicitly. In [57] he gave a somewhat different derivation of the squared character values. Srinivasan [53] resolved the sign question in a reasonable way, but only under the unpalatable assumption that \( q \) is “sufficiently large”. Ultimately an elegant formulation of the signs was developed, following suggestions of Springer: see [13 and 25]. It turns out that the sign of \( \text{St}(s) \) depends just on the parity of the “semisimple ranks” of the reductive groups \( \mathcal{G} \) and \( C_{\mathcal{G}}(s) \) relative to the finite field of definition. So in principle \( \text{St} \) is an easy character to compute.

1.3 Alternating sum formula for \( \text{St} \). In 1965 Curtis [19] (and independently, Feit) discovered a remarkable formula expressing \( \text{St} \) as a \( \mathbb{Z} \)-linear combination of induced characters. This brings into play not just \( B \), but all the subgroups \( P \) of \( G \) containing \( B \) (these are called, for somewhat arcane reasons, “parabolic subgroups”). These subgroups correspond in a natural way to the \( 2^r \) subsets \( I \) of the set \( S \) of “simple” reflections in \( W \), where \( r = \dim \mathcal{G} \) is the rank of \( \mathcal{G} \). So we may denote them by \( P_I \), with \( B \) corresponding to the empty set and \( G \) to the whole set. Write \( 1^G_P \) for the character induced from the trivial character of \( P \). Then the formula of Curtis reads

\[
\text{St} = \sum (-1)^{|I|} 1^G_{P_I} \quad \text{(sum over } I \subseteq S). 
\]

In developing his formula, Curtis also characterized \( \text{St} \) as the unique irreducible constituent of \( 1^G_B \) which fails to occur in any other \( 1^G_P \) (when \( P \neq B \)). The proof of the formula depends on an analogous formula for the alternating character \( \varepsilon \) of \( W \), discovered by Solomon in his work on the orders of Chevalley groups. Each set \( I \) of simple reflections in \( W \) generates a “parabolic” subgroup \( W_I \), and Solomon showed that

\[
\varepsilon = \sum (-1)^{|I|} 1^W_{W_I}.
\]

All of this can be developed in the slightly more general setting of finite groups with \( (B, N) \)-pair, by taking the alternating sum formula as the definition of \( \text{St} \), and showing that it agrees with Steinberg’s original construction in the case of groups of Lie type. In sections (6.1)–(6.4) of Carter’s book [13] this viewpoint is developed systematically. One advantage of this approach is that it lends itself to inductive calculation of the actual character values of \( \text{St} \): Each parabolic subgroup has a “Levi decomposition” \( P = L V \) (corresponding to reductive times unipotent radical in algebraic groups), and for the Levi subgroup \( L \) one can define a Steinberg character in the same spirit as for \( G \).
Then it turns out that restricting $S$ to $P$ is the same thing as inducing the Steinberg character of $L$ up to $P$.

1.4 Induced characters and Hecke algebras. The study of $S$ as a constituent of the induced character $1_{G}^{\mathcal{G}}$ brings out nicely its deeper connection with the alternating character $\varepsilon$ of $W$. To find the irreducible constituents of an induced character, it is natural to look at the “Hecke algebra” consisting of $CG$-endomorphisms of the induced module in question. The irreducible representations of the Hecke algebra are in natural bijection with the sought-for irreducible constituents, with the degree of the former measuring the multiplicity of the latter as a summand of the induced module.

In the case at hand, the Hecke algebra turns out to have a remarkable resemblance to the group algebra $CW$. Both algebras can be viewed as specializations of a “generic” algebra, in such a way that the representations of the Hecke algebra correspond naturally to representations of $W$. Moreover, the Hecke algebra has canonical generators in bijection with the simple reflections in $W$, so a representation is determined by its effect on these generators. Putting all the pieces together, the alternating character $\varepsilon$ of $W$ corresponds to the 1-dimensional representation of the Hecke algebra sending all generators to $-1$, and this in turn corresponds to the constituent $S$ of $1_{G}^{\mathcal{G}}$. This is due to Curtis-Iwahori-Kilmoyer [23], and is surveyed in Curtis [21], while a very comprehensive exposition is given in Carter [13, Chapter 10].

1.5 Homological construction of $S$. The alternating sum formula for the Steinberg character in (1.3) suggested to Solomon that the representation should have a homological origin for which this would be an Euler character. He outlined such a construction in [52] (with advice from Tits); later this was refined and generalized by Curtis-Lehrer [24] (cf. [22]), and by Curtis-Lehrer-Tits [25]. In particular, the version in [25] provides an elegant computation of all character values of $S$.

Solomon's basic idea stems from an analogous construction of the alternating character $\varepsilon$ of $W$, using the “Coxeter complex” $\Gamma$ of $W$, based on the collection of parabolic subgroups, whose left cosets are taken as simplexes. The geometric realization of $\Gamma$ is a sphere, and $W$ acts on the homology: the 0-homology affords the trivial character, while the top homology affords $\varepsilon$. From the Hopf trace formula one derives the alternating sum formula for $\varepsilon$ in (1.3). The appropriate geometric structure in the case of $G$ is the Tits complex (or building) $\Delta$, based on the collection of parabolic subgroups of $G$ (those containing a conjugate of $B$). The complex $\Delta$ has the homotopy type of a bouquet of $(r-1)$-spheres, $r = \text{rank}$ as before. (In the case $r = 1$, one has to argue separately.) The integral homology group $H_{r-1}(\Delta)$ is free abelian of rank $q^m$ and, after tensoring with $\mathbb{Q}$, affords the Steinberg representation. Again one can apply the Hopf trace formula to obtain the alternating expression of (1.3).

The main improvement in [25] is the formulation in functorial terms of the Tits building of a group $G$ coming from a reductive (not necessarily semisimple) algebraic group, which facilitates induction on semisimple rank: the Levi subgroup of a parabolic subgroup of $G$ always comes from a reductive group. Roughly speaking, if $x = su$ ($s$ $p$-regular, $u$ $p$-singular, $su = us$), the fixed point set of $x$ on the building of $G$ is the same as the fixed point set of $u$. 
on the building of the centralizer of $s$. This feeds into the Hopf trace formula and yields a computation of the character values of $\text{St}$.

1.6 **An application: counting unipotent elements.** Steinberg [57] (cf. [13, 6.6]) devised a clever way to count the unipotent elements in a finite group of Lie type, by exploiting the values of the Steinberg character. For a group $G$ of the sort we are considering, the answer is surprisingly simple: $G$ has $q^{2m}$ unipotents. The idea of the proof can be sketched fairly quickly.

The Jordan-Chevalley decomposition (semisimple times unipotent) in $G$ can be carried out in $G$, where it coincides with the sort of decomposition possible in any finite group: $x \in G$ can be written uniquely as $x = su$, where $s$ is $p$-regular, $u$ is $p$-singular, and $su = us$. For fixed $s$, the number of distinct choices for $u$ is $P(s) = \text{number of } p\text{-singular elements in } C_G(s)\text{—just the number we are looking for, if } s \in Z(G)$. Thus $|G| = \sum P(s)$, summed over the $p$-regular elements.

If we set $Q(s) = \text{order of } \text{Sylow } p\text{-subgroup of } C_G(s)$, we have $|\text{St}(s)| = Q(s)$ for $p$-regular $s$. In particular, $Q(s) = q^m$ if $s \in Z(G)$. On the other hand, $\text{St}(x) = 0$ if $p$ divides the order of $x$. The orthogonality relations for $G$ (applied to the pairing of $\text{St}$ with itself) then yield: $|G| = \sum Q(s)^2$, summed over the $p$-regular elements. The resulting equation is the key to the counting argument:

$$\sum P(s) = \sum Q(s)^2.$$ 

What has to be shown is that $P(s) = Q(s)^2$ if $s \in Z(G)$. The proof proceeds by induction on the dimension of the (semisimple) group $G$, taking into account the fact that the derived group of the centralizer of a semisimple element is a semisimple group containing all the unipotents in that centralizer. So if $s \notin Z(G)$, induction will show that $P(s) = Q(s)^2$, and all these terms can be omitted from the equation above. Then we can divide the remaining equation by the order of $Z(G)$ to complete the proof.

1.7 **Role of St in Deligne-Lusztig theory.** So far we have looked at the Steinberg character as an isolated irreducible character of $G$. How does it fit into the character table as a whole? For groups like $\text{SL}_n(q)$, it immediately stands out from the crowd by being the sole character of degree $q^m$ (indeed, no other character has degree divisible by so large a power of $p$). But what interactions are there between $\text{St}$ and other characters? Multiplying by $\text{St}$ turns out to be interesting, as we shall see in a moment (and in the context of projective modules in §3).

The role of $\text{St}$ is most apparent in the wider context of Deligne-Lusztig characters [26], the main subject of Carter's exposition [13]. Here we broaden for a moment our use of the symbol $T$. Deligne and Lusztig assigned to each (not necessarily split-type) “torus” $T$ in $G$ and to each $\mathbb{C}$-valued character $\theta$ of $T$ a virtual character $R_{T,\theta}$ of $G$ ($\mathbb{Z}$-linear combination of irreducible characters), coming from a cohomological construction. Most of these virtual characters are in fact irreducible characters (up to sign), and each irreducible character occurs as a constituent of some $R_{T,\theta}$. When $T$ is of split type, $R_{T,\theta}$ is just the induced character obtained by viewing $\theta$ as a character of $B$ and then
inducing up to \( G \). In particular, \( \text{St} \) occurs as a constituent when \( \theta = 1 \). But \( \text{St} \) may also occur for other tori, with \( \theta \) again being the trivial character. All irreducible constituents arising from the use of \( \theta = 1 \) are labeled \textit{unipotent characters}. They play a fundamental role in the working out of the Deligne-Lusztig theory. \( \text{St} \) may be viewed as a “regular” unipotent character in the rough analogy which exists between characters and conjugacy classes.

\( \text{St} \) itself contributes at an early stage of the theory to the calculation of the degree of \( R_{T,\theta} \) or, more generally, to the calculation of the character values at semisimple elements of \( G \). This comes about through the fact that the product of \( R_{T,\theta} \) and \( \text{St} \) is (up to a well-determined sign) simply the induced character \( \theta_G^G \). Since the values of \( \text{St} \) at semisimple elements are known (and nonzero!), the values of \( R_{T,\theta} \) can be found. In particular, the degree of \( R_{T,\theta} \) is (up to sign) \(|G|/q^m|T| \). It is, however, a far more subtle matter to determine the character values at non-semisimple elements.

1.8 \textit{Analogues of St}. In “semisimple” Lie theory one frequently notices parallel developments going on, even when the subjects are logically independent: one may be dealing with real or complex Lie groups, \( p \)-adic groups, algebraic groups over fields of prime characteristic, Lie algebras over various fields, or finite groups of Lie type. This tendency is especially clear in representation theory.

In the case of the Steinberg representation, there is a striking parallel in the study of (semisimple or reductive) \( p \)-adic groups such as \( \text{SL}_n(Q_p) \). Here a particular type of square-integrable irreducible admissible representation, called “special”, appeared in the work of Matsumoto [47] and Shalika [51] (following the study of the rank one case by Gel’fand and his co-workers). It was soon clear that this representation has a precise formal resemblance to the Steinberg representation, cf. Borel [10], Borel-Serre [11], Casselman [15], Garland [30], Harish-Chandra [33]. For example, the “character” has an alternating sum expression like the one for \( \text{St} \), and the representation itself corresponds naturally to a “Hecke algebra” character taking value \(-1\) at the canonical generators. Moreover, there is a Bruhat-Tits “building” naturally associated with the \( p \)-adic group, whose top-dimensional homology (with compact supports) has natural \( G \)-action affording the special representation.

In another direction, Lees [45] found a representation of \( \text{GL}_n(Z/p^hZ) \) with some properties similar to those of \( \text{St} \) in the case \( h = 1 \). For example, its character values resemble those of \( \text{St} \), and it occurs as a large constituent of the representation induced from the trivial representation of a “Borel subgroup” (though it is not usually irreducible). In unpublished work, he also found a homology realization of the largest irreducible constituent of this representation.

2. \( \text{St} \) as an irreducible modular representation.

2.1 \textit{Reduction modulo} \( p \). Because the degree of the Steinberg representation is divisible by (indeed equal to) the full power of \( p \) dividing the order of \( G \), the Brauer-Nesbitt Theorem implies that it remains irreducible after “reduction modulo \( p \)”. (This also follows more directly from Steinberg’s original construction outlined above.) Since the representing matrices can be chosen (relative to
a suitable basis) to have ordinary integers as entries, it is clear what it means to reduce modulo \( p \): one simply reduces all matrix entries modulo \( p \), obtaining a representation of \( G \) over \( \mathbb{F}_p \) or its algebraic closure \( K \).

In general, the process of reduction modulo \( p \) is somewhat more complicated to describe. Moreover, the original choice of basis can have a serious effect on the nature of the modular representation obtained: a different choice could lead to an inequivalent representation, although the composition factors (with multiplicity) are always the same. Because \( St \) remains irreducible, none of these subtle points arise. As a matter of notation, we shall continue to use \( St \) to denote this representation after reduction modulo \( p \), relying on the context to avoid confusion.

There is an obvious question at this point: *Where does the modular representation \( St \) fit into the overall picture of irreducible \( KG \)-modules?* It follows from Brauer theory that the number of nonisomorphic irreducible \( KG \)-modules is equal to the number of \( p \)-regular conjugacy classes. So this collection of modules is somewhat smaller than the corresponding collection of irreducible \( CG \)-modules. Moreover, the dimensions of irreducible modules in characteristic \( p \) need not divide the order of \( G \). (In this, as in many other ways, \( St \) is exceptional.)

From work of Steinberg [56] we have a clearcut computation of the number of \( p \)-regular classes of \( G \) when \( G \) is simply connected: the number is \( q^r \), if \( G = \mathfrak{G}(\mathbb{F}_q) \) and \( r = \text{rank } G \). Because the formulation is more complicated for other types of groups, we shall stick with the simply connected case here.

2.2 The Curtis-Richen approach. Before describing Steinberg's work on modular representations in more detail, we should make a few remarks about the later development by Curtis [20] and Richen [50] of a more self-contained development in the framework of "split" \((B, N)\)-pairs, cf. also [14]. From the viewpoint of a finite group theorist, it is desirable to understand the irreducible modular representations of \( G \) in the context of finite groups alone. The formalism of the Tits system in fact permits a reasonably satisfactory treatment—up to a point! But so far no one has been able to get good information about degrees or characters of representations without invoking the algebraic groups.

What Curtis and Richen accomplish is a self-contained classification and construction of irreducible \( KG \)-modules, though the term "construction" has to be qualified somewhat: as noted above, one doesn't get full information about dimensions or characters. In spite of this serious limitation, their approach does have the advantage of directness. It also applies to a somewhat wider class of groups, but we will stick with \( G \).

Taking their cue from the algebraic group case (to be discussed below), Curtis and Richen classify irreducible \( KG \)-modules by "weights" which involve especially the characters of the torus \( T \). The unpleasant complication here is that the \( T \)-action alone may fail to distinguish one module from another, so the more detailed actions of root groups (or parabolic subgroups) have to be examined carefully—for example, to distinguish the Steinberg module from the trivial module. The problem is solved neatly by attaching to a suitable character of \( T \) a string of \( r \) numbers with value 0 or -1. The trivial character
of $T$ with a string of 0's identifies the trivial $KG$-module. Replacing each 0 by $-1$ yields St.

Without attempting a more technical description, we should point out that Curtis and Richen achieve an indirect count of the number of $p$-regular conjugacy classes by constructing and counting the irreducible $KG$-modules. They actually construct the modules as minimal left ideals of the group algebra $KG$ generated by "weight elements". If $m$ is such an element, $KGM$ turns out to equal $KVm$ (where $V$ is a Sylow $p$-subgroup "opposite" to $U$). This yields a rough upper bound on the dimensions of irreducibles, namely $q^m$. In the case we are considering (where $G$ is equal to its derived group), they arrive at a unique module of the maximum possible dimension $q^m$, which must therefore be St.

Thus our question in (2.1) may be answered in part by pointing out that St is the unique largest irreducible $KG$-module.

2.3 Representations of the algebraic group. The original classification of irreducible $KG$-modules was a byproduct of the classification of irreducible "rational" representations of the algebraic group $G$. The ideas here are due to Chevalley, generalizing the classical "highest-weight" classification of Cartan and Weyl. Without going into too many details, we can indicate the methods used. The idea is to exploit the completely reducible action of the algebraic torus $T$ on a rational $G$-module $V$. Here $V$ breaks up into a direct sum of "weight spaces" $V_\mu$, on which $T$ acts via some character $\mu$: $T \rightarrow K^*$. The characters form a free abelian group $X(T)$, usually written additively, on which the choice of Borel subgroup (or positive roots) imposes a natural partial ordering. Relative to this ordering, an irreducible $G$-module has a unique highest weight $\lambda$, which is a "dominant" weight. Since $G$ is assumed to be of simply connected type, there is in fact a bijection between dominant weights $\lambda$ and irreducible modules $L(\lambda)$. (See [9 or 37] for further discussion.)

The actual construction of irreducible modules is far from being straightforward; indeed, the dimensions and "formal characters" (recording all weight space dimensions) are still unknown in general. But Lusztig's well-known 1979 conjecture offers substantial hope of answering such questions—at least when $p$ is not too small. One can extract $L(\lambda)$ as a quotient of a "Weyl module" $V(\lambda)$, which comes by reduction modulo $p$ from a well-chosen $\mathbb{Z}$-form of an irreducible module of similar highest weight for the corresponding complex semisimple algebraic group, cf. [9]. This gives only a crude upper bound for the dimension of $L(\lambda)$, which is known to be usually much smaller than $V(\lambda)$.

There is one fortunate simplification here, the Steinberg Tensor Product Theorem [56] (cf. [17]), which had been observed already by Brauer-Nesbitt in the very special situation of $SL_2(K)$. The free abelian group $X(T)$ has a canonical basis consisting of "fundamental dominant weights" $\omega_1, \ldots, \omega_r$. The dominant weights are just the nonnegative linear combinations $\sum c_i \omega_i$. Call such a weight $\lambda$ "restricted" if its coefficients lie in the interval $[0, p-1]$. Then an arbitrary dominant $\lambda$ has an expansion $\lambda = \lambda_0 + p\lambda_1 + \cdots + p^s\lambda_s$, with all $\lambda_j$ restricted. Steinberg showed that $L(\lambda)$ is isomorphic to $L(\lambda_0) \otimes L(p\lambda_1) \otimes \cdots \otimes L(p^s\lambda_s)$, where $L(p^s\lambda_s)$ is itself isomorphic to a "twisted" version of $L(\lambda_s)$, involving $p^s$th powers of matrix entries. So the study of
irreducible $G$-modules boils down to the study of the finitely many ones with restricted highest weights.

2.4 Restriction to the finite groups. In the fundamental paper [56], Steinberg proved that the irreducible $KG$-modules can all be obtained by restricting certain irreducible rational $G$-modules $L(\lambda)$ to $G$. (See [9] for a somewhat different presentation of the ideas.) For example, when $G$ is a split (Chevalley) group such as $SL_n(q)$, defined over a field of $q = p^t$ elements, one gets all the (nonisomorphic) irreducible $KG$-modules by restricting those $L(\lambda)$ for which the coordinates of $\lambda$ lie in the closed interval $[0, q]$. Note that the total number of such modules is $q^r$, in conformity with (2.1) above.

Where does St fit into this classification of irreducible $KG$-modules? Recall from (2.3) that $\dim L(\lambda)$ is bounded above by $\dim V(\lambda)$, which in turn is given by Weyl's classical dimension formula. It is easy to check that $\dim V(\lambda) < \dim St$ for all allowed weights $\lambda$ except $(q-1)p$, where $\rho = \sum \omega_i$, whereas $\dim V((q-1)p) = \dim St$. From this we conclude immediately that $V((q-1)p)$ is equal to $L((q-1)p)$ and restricts to the Steinberg representation of $G$. Note that St may now be regarded as essentially a representation of $G$, which in fact comes by reduction modulo $p$ from a characteristic 0 representation of an algebraic group (or Lie algebra). This origin of St in characteristic 0 is quite independent of the finite group setting discussed in the first part of this article!

It is also interesting to remark that St for $G$ arises as a twisted tensor product (2.3) involving just the Steinberg module for the corresponding group over the prime field $F_p$. To see this, write down the $p$-adic expansion of the weight $(q-1)p$.

One further note: From the viewpoint of the algebraic group $G$, it is easy to see that St is self-dual (essentially because the coefficients of the highest weight are so symmetric). This could also be worked out in the Curtis-Richen approach (2.2), and is of course an obvious consequence (for the type of group $G$ we are allowing) of the fact that St is the only irreducible $KG$-module of its dimension.

2.5 Representations of the Lie algebra. There is yet another angle from which the Steinberg representation can be viewed, involving the Lie algebra $g$ of $G$. In the earlier work of Curtis, as well as in [56], representations of $g$ play a major role. There is a natural “differentiation” process which allows us to regard rational representations of $G$ as “restricted” representations of $g$ (compatible with the natural $p$th power operation in $g$ that makes it a restricted Lie algebra). To sum up briefly what is true: The $p'$ modules $L(\lambda)$ for restricted $\lambda$ yield precisely the irreducible restricted representations of $g$, and St is the largest of these. So there is a nice parallel between $g$ and $KG$ for $G = G(F_p)$.

The parallel is even nicer if we recall that restricted representations of $g$ are essentially the same as $u$-modules, where $u$ is the “restricted enveloping algebra” (an associative algebra of dimension $p^{\dim g}$). When $p$ is large, the dimensions of $u$ and $KG$ are of the same order of magnitude, and both algebras are “symmetric” finite dimensional Hopf algebras. So the resemblance is in some ways quite strong.
The fact that $V((p - 1)p)$ remains irreducible on restriction to $g$ can be understood in fairly direct ways, as remarked already in [56]. The development of the "linkage principle" (cf. [2, 34, 40]) makes the situation transparent: all composition factors of a Weyl module $V(\lambda)$ other than $L(\lambda)$ must have "linked" highest weights lying below $\lambda$, and $(p - 1)p$ is linked only to itself. In effect, $(p - 1)p$ is maximally isolated among the restricted weights.

One side remark: Much of the theory just sketched carries over step-by-step to the (infinite dimensional) affine Kac-Moody Lie algebras and their "highest-weight" representations. For example, there is a suitable $Z$-form in the characteristic 0 enveloping algebra of such a Lie algebra which enables one to reduce modulo $p$. The analogue of the modular Steinberg representation shows up here in an interesting, but somewhat mysterious light, as explored by Arnon [4]. Arnon discovers in particular a nice connection with the $p$-adic "special representation" mentioned in (1.8) as a natural analogue of the characteristic 0 Steinberg representation.

2.6 Representations of hyperalgebras. To shift the viewpoint again, $u$ corresponds (as a Hopf algebra) to an "infinitesimal subgroup" $\mathcal{G}_i$ of $\mathcal{G}$, the kernel of the Frobenius ($p$th power) endomorphism. The differentiation process amounts to restricting representations from $\mathcal{G}$ to $\mathcal{G}_i$, which is a normal subgroup scheme. The analogue of Clifford's Theorem for finite groups can be developed here, making for an especially nice interpretation of Steinberg's tensor product theorem discussed above, cf. [17].

The parallel between $u$ and the group algebra of $\mathcal{G}(F_p)$ actually has higher analogues, explored for example in [16, 36, 38]. Corresponding to the group $G$ over a field of $p^n$ elements, there is an analogous "hyperalgebra" $u_n$ (or infinitesimal group $\mathcal{G}_n$) whose irreducible representations correspond precisely to those of $G$. In particular, the Steinberg module $L((p^n - 1)p)$ is the largest module here.

2.7 Induced modules and sheaf cohomology. The study of irreducible $\mathcal{F}$-modules $L(\lambda)$ as quotients of Weyl modules $V(\lambda)$ as in [9] has the advantage of concreteness. The dimensions of weight spaces (i.e., the $\mathcal{F}$-module structure) come from the classical Weyl theory, facilitating the study of how things restrict to $G$. However, this approach is somewhat unsatisfying from a theoretical point of view, since it does not construct the modules $V(\lambda)$ directly as characteristic $p$ objects. There turns out to be a dual approach which is intrinsic and useful.

Briefly put, each character $\lambda$ of $\mathcal{F}$ extends to a 1-dimensional representation of the Borel group $\mathcal{B}$ (trivial on the unipotent radical $U$), and then there is a reasonable notion of "induction" from rational $\mathcal{F}$-modules to rational $\mathcal{F}$-modules (explored especially by Cline-Parshall-Scott). The resulting $\mathcal{F}$-module may be denoted $H^0(\lambda)$, in view of the alternative interpretation as the 0th sheaf cohomology of $G/B$ relative to the line bundle associated with $\lambda$. (Here $\mathcal{F}$ acts in a natural way on global sections.) Now $L(\lambda)$ occurs as the unique irreducible submodule of $H^0(\lambda)$. This interpretation of induction leads in some interesting directions, studied especially by Andersen. One starts with Kempf's vanishing theorem [41] (cf. [3, 32]), which asserts (for $\lambda$ dominant) the vanishing of all higher sheaf cohomology. Then one observes that the Euler
characteristic is still what it was in characteristic 0, given by Weyl's character formula (cf. [28] for a self-contained derivation of this fact). It follows readily that \( H^0(\lambda) \) is dual to a Weyl module \( V(\lambda^*) \), where \( L(\lambda^*) \) is dual to \( L(\lambda) \).

In looking at some of the cited literature (such as [42]), the reader will be struck with the ubiquitous presence of the Steinberg representation in this setting. In part this is a result of the convenient fact that the line bundle associated with \((q - 1)p\) is ample, which already yields some weak vanishing theorems of the type Kempf was after. The proofs of Kempf's theorem by Andersen [3] and Haboush [32] give an especially prominent role to \( St \), which is equally true of the proof of Mumford's conjecture by Haboush [31] (cf. also the version in [36] and the treatment in [38, II.10]).

3. \( St \) as a projective module.

3.1 Principal indecomposable modules. The Brauer-Nesbitt Theorem asserts not only that \( St \) remains irreducible upon reduction modulo \( p \), but also that it is a principal indecomposable module (abbreviated PIM) for \( KG \). Let us recall what this means. When \( KG \) is decomposed as a direct sum of indecomposable left ideals, the summands are PIM's: indecomposable projective modules, in modern terminology. Each of these has a unique irreducible quotient \( L(\lambda) \), so the PIM is characterized as the "projective cover" of \( L(\lambda) \) and may be denoted \( U(\lambda) \). In fact, the algebra \( KG \) is "symmetric," implying that \( U(\lambda) \) is also an injective \( KG \)-module with the unique irreducible submodule \( L(\lambda) \). It is also a general fact that the order of a Sylow \( p \)-subgroup of \( G \) divides the dimension of each PIM. Thus \( St \) is as small as a PIM can possibly be, and no other irreducible module \( L(\lambda) \) can be simultaneously projective—always under the simplifying hypothesis that \( G \) is of universal type, hence equal to its derived group.

The Brauer-Nesbitt theorem is usually quoted in the context of block theory: The indecomposable two-sided ideals of \( KG \) are called blocks and have a nice description in terms of PIM's. Namely, sort the PIM's into equivalence classes by requiring that \( U(\lambda) \) be in the same class as \( U(\mu) \) if they share at least one composition factor. Then the sum of all PIM's in one class is a block of \( KG \). It is clear that an irreducible PIM such as \( St \) will lie in a block involving no other PIM's; by general principles, this block is a direct sum of \( q^m \) copies of \( St \). Such a block is said to have "defect 0" and is the simplest type of block in Brauer's theory. For a simple group of Lie type, a theorem of the author says that \( KG \) has precisely one other block, involving all PIM's except \( St \). (More generally, if \( G \) has a nontrivial center, the order of the center is the number of other blocks.)

3.2 Tensoring with \( St \). For any finite group, the tensor product of a projective module with an arbitrary module is again projective. (This can be proved for any finite dimensional Hopf algebra, not just a group algebra.) In our context, the Steinberg module plays a pivotal role—as the largest irreducible module and smallest projective module for \( KG \). Initially, \( St \) is the only PIM known explicitly. By tensoring it with various irreducible modules \( L(\lambda) \), one can hope to get hold of many other PIM's. This is a key idea in the early computations of Cartan invariants (composition factor multiplicities of PIM's) by Feit and
his students Chastkofsky and Cheng. The only catch is that $L(\lambda) \otimes St$ may be very large, hence may have many PIM's as direct summands.

The first group studied in detail was $SL_2(q)$. Jeyakumar [39] showed how to sort out the PIM's effectively, by taking advantage of the $G$-module structure on the above tensor products. His rather computational arguments were streamlined by the author [35], who also made some progress on the general case. The first really general results were obtained by Ballard in his 1974 thesis [5] (written up in [6] and improved in [7]). Ballard showed that any PIM $U(\lambda)$ for $KG$ can be obtained as a distinguished direct summand of the tensor product of $St$ with an irreducible module $L(\mu)$ correlated naturally with $\lambda$. As a byproduct of the argument, one sees that the "Brauer character" of a PIM is a $\mathbb{Z}$-linear combination of products of the Steinberg character with $W$-orbit sums. However, one does not yet get explicit formulas for the characters and dimensions.

Ballard showed in effect that $St$ "divides" all other projectives. This makes sense if one works in the Grothendieck ring of $KG$-modules, where direct sum and tensor product provide the ring operations. In this ring the projective modules generate an ideal, which in turn is generated by $St$. Lusztig [46] developed this viewpoint (for a wider class of groups than we are considering). Afterwards Alperin [1] and Feit [29] analyzed further what is going on here, e.g., what special features of $St$ permit it to divide all projectives.

3.3 The infinitesimal parallel. It was pointed out in §2 that the irreducible $G$-modules $L(\lambda)$ having "restricted" highest weight yield the irreducible modules both for $KG$, $G = G(F_p)$, and for $u$, the restricted universal enveloping algebra of the Lie algebra $g$. A similar parallel occurs for the field of $p^n$ elements if we introduce hyperalgebras $u_n$. There is an analogous—but more complicated—pattern for projective modules. It begins with the observation that $St$ is a PIM for $u_n$, when $St$ is the $G$-module of highest weight $(p^n - 1)p$. If $n = 1$, this appears first in the 1963 thesis of Nielsen [49]. In his Lie algebra set-up (requiring $p > 7$), he was able to give an explicit set of generators for the minimal (right) ideals in $u$ and to show that only one such ideal (namely $St$) fails to be nilpotent and hence yields a PIM. Other arguments, using the linkage principle, were later given by the author in the general case of $u_n$, cf. [36].

As we mentioned above, it is a general fact about finite dimensional Hopf algebras that tensoring with a projective module produces another projective module. This again makes $St$ a natural tool in the study of PIM's of $u_n$. But the fact that $St$ is a $G$-module gives even more leverage. The arguments of Jeyakumar [39] in the case of $SL_2$ suggested to the author and to Verma the possibility of "lifting" PIM's of $u$ to $G$-modules in all cases; these would in turn restrict to projective $KG$-modules, having been constructed as direct summands in tensor products with $St$—and would be usually (but not always!) indecomposable for $KG$. Ballard [7] legitimized this procedure, for $p$ not too small, but it has not yet been proved for arbitrary $p$. (Jantzen did succeed in lowering Ballard's bound on $p$ somewhat.) Independently, Chastkofsky and Jantzen then found an explicit formula for the way a lifted PIM of $u$ (or $u_n$) would decompose into PIM's for $KG$, based on the prior knowledge of the irreducible modules $L(\lambda)$. This indirect route to the PIM's of $KG$ has in fact
been the source of most of our detailed knowledge to date, e.g., multiplicities of composition factors of PIM’s (Cartan invariants).

The technique of tensoring with St has been fruitful in a number of questions. For example, Benson [8] has worked out the Loewy structure of PIM’s for $\text{Sp}_4(3)$ and related groups by exploiting heavily the representation theory of the corresponding algebraic group in the context of tensoring with St. Other examples occur in the work of Andersen [3] and Haboush [31, 32] cited earlier.

### 3.4 Injectives and blocks for the algebraic group

The algebraic group $G$ serves as a bridge between the finite subgroups $G = G(F_q)$, $q = p^n$, and the infinitesimal groups $G_n$ (or associated hyperalgebras $u_n$). Irreducible modules in each case are obtained simply by restricting finitely many of the irreducible rational $G$-modules $L(\lambda)$, and projective/injective modules (at least for $p$ not too small) are obtained by restricting certain indecomposable rational $G$-modules $Q(\lambda)$. In the latter case, however, $Q(\lambda)$ may fail to be indecomposable for $KG$, in which case the PIM $U(\lambda)$ has to be extracted as a direct summand. Apart from the uncertainty about small primes, the modules $Q(\lambda)$ are a bit mysterious. They are definitely not injective in the category of rational $G$-modules. Indeed, this category has to be defined to admit infinite dimensional (but locally finite dimensional) modules in order to have enough injectives—and it does not have projectives. The injective hull $I(\lambda)$ of $L(\lambda)$ exists, but is always infinite dimensional, cf. [16].

The pieces do fit together, however (at least when $p$ is not too small): $I(\lambda)$ is just the direct limit of finite dimensional modules $Q(\lambda)$, where $\lambda$ is fixed but $Q(\lambda)$ gets larger as $n$ gets larger—and should properly be denoted as $Q(n, \lambda)$ when defined as a PIM of $u_n$. Such a theorem was first obtained by Ballard [7], with later improvements by Donkin and by Jantzen.

While St itself is not an injective module for $G$, it does play some special role in the description by Donkin [27] of “blocks” of $G$. Here the Steinberg modules for $u_n$ ($n = 1, 2, 3 \ldots$) all lie in distinct blocks.

### Conclusion

Besides showing how the Steinberg representation exemplifies the Brauer-Nesbitt Theorem, this survey should help to explain its ubiquity in the representation theory (and related cohomology) of groups of Lie type. It is indeed a special representation.

### References


