Introduction

1

This book is about the representation theory of analytic groups (an analytic group is a connected complex Lie group and a representation of it as matrices of size $n \times n$ is a holomorphic homomorphism from the analytic group to the group $GL_n\mathbb{C}$ of all invertible $n \times n$ complex matrices). As it is usually viewed, the main problem of representation theory is: given the group, determine, in terms of some 'parameters', the representations of the given group. For example, the classification of representations of a simply-connected simple Lie group in terms of the high weights of irreducible components, and the description of the possible high weights from the root system of the Lie algebra of the group, is a profound and inspiring solution to the problem for the groups to which it applies. (One can get an idea of how inspiring this solution was by consulting, for example, the bibliography of [1]†.)

There is also, however, the converse problem, which will be our major concern: given its representations, determine the group. The problem, as stated, is not very well-posed (what does it mean to be 'given the representations'?) although, as is often the case in the development of mathematics, that was not a serious deterrent historically (the historical development is summarized below), and currently the concepts of category theory allow a precise statement, as we shall see. It turns out, however, that the problem does not always have a solution. There are analytic groups that are not isomorphic, and yet their categories of representations are the same; examples are given in Chapter 1, Example E. Thus our problem becomes: given the representations of an analytic group, what do we know?

† Bracketed references refer to the bibliography.

As we shall see, we know quite a bit. To explain the answer, we need to be precise now about the problem. By the familiar mechanisms, we can regard a representation of size $n \times n$ of the analytic group G as giving a structure of G-module on an n-dimensional complex vector space, and vice-versa; see (1.1), (1.2) below.† But by dealing with the G-modules, we can use the additional idea of a G-module homomorphism (1.4) to conceive of the category $\text{Mod}(G)$ of (finite-dimensional, analytic) G-modules. Thus, we interpret the phrase 'given the representations' of G to mean that the category $\text{Mod}(G)$ is known to us, in the sense that it is a specified subcategory of the category of finite-dimensional complex vector spaces. We will also need to know that it is closed under the vector space operations of tensor product, direct sum, and linear dual. This category, while it does not determine G, does determine an (abstract) group, which we denote $\text{Aut}_{\otimes}(\text{Mod}(G))$, and an algebra $R(\text{Mod}(G))$ of complex functions on it such that the pair $(\text{Aut}_{\otimes}(\text{Mod}(G)), R(\text{Mod}(G))$ completely characterizes the category $\text{Mod}(G)$. Thus, we can conclude: given the representations of G, we know $\text{Aut}_{\otimes}(\text{Mod}(G))$ and $R(\text{Mod}(G))$, and conversely.

Ths neatly answers the question of what one knows when one knows the representations of an analytic group, but does not yet speak of the original problem of recovering G from its representations. It turns out that there is a group homomorphism $G \to \text{Aut}_{\otimes}(\text{Mod}(G))$, which we also use to make $R(\text{Mod}(G))$ an algebra of functions on G, and using these we elucidate the structure of $\text{Aut}_{\otimes}(\text{Mod}(G))$ and $R(\text{Mod}(G))$.

The relative logical positions of these two operational modalities (obtaining $\text{Aut}_{\otimes}(\text{Mod}(G))$ from $\text{Mod}(G)$ as a category, and studying $\text{Aut}_{\otimes}(\text{Mod}(G))$ via the homomorphism from G to it) need to be kept clear: in the first we are not permitted to glimpse G, but only its category $\text{Mod}(G)$ of representations, while in the second we employ G directly, assuming it known in order to describe its associated group $\text{Aut}_{\otimes}(\text{Mod}(G))$. The opportunities for confusion

†Results, remarks and definitions are so numbered that the digit to the left of the decimal indicates the chapter in which they occur.

are obvious, especially since sometimes it will be necessary to operate in both modes simultaneously, but our general trend is to be as follows: as much as possible, the first will be done first – mostly in Chapter 2 – while the second will be done subsequently.

A detailed summary of the book's contents is given in the third section of this introduction. The second section contains a brief history and the fourth contains some general comments about assumed prerequisites.

2

We can begin the story of recovering groups from their representations with Pontryagin's duality theorem of 1934 [26]: here G is a compact abelian group, $\hat{G}$ is the dual discrete group of continuous (complex) characters of G, and $\hat{\hat{G}}$ is the dual group of complex characters of $\hat{G}$. Then his 'first fundamental theorem' asserts that G and $\hat{\hat{G}}$ are isomorphic. A bit of translation is needed to see this theorem in the context we introduced above. First, finite-dimensional representations of G are direct sums of one-dimensional representations $G \to GL_1\mathbb{C}$ which are the characters of G; G-homomorphisms between these one-dimensional G-modules are given by scalar multiplications only, so can be ignored. But the tensor product of these G-modules is what determines the composition in $\hat{G}$. Thus we can regard knowledge of $\hat{G}$ and knowledge of Mod(G) equivalent in the compact abelian case. The recovery procedure here also deserves some comment: the idea is to define complex functions on $\hat{G}$ (the set of one-dimensional representations of G) which respect the composition rule in $\hat{G}$; that these turn out to be characters of $\hat{G}$ should be regarded as a consequence of the fact that only one-dimensional representations are involved.

This description of the recovery procedure is given with the benefit of hindsight. Once stated this way, the idea of how to extend Pontryagin's procedure to nonabelian compact groups is clearer: again, one can stick to the set $\hat{G}$ of representations of the compact group G, which is now to have the composition rules of direct sum and tensor product of representations, and now $\hat{\hat{G}}$ is to be made of 'representations' of $\hat{G}$, that is, matrix valued functions respecting the composition rules. In this context, Tannaka proved his duality theorem of 1938 [29] which states that G and $\hat{\hat{G}}$ as above are isomorphic.

Chevalley's 1946 book [3] includes a proof of Tannaka's duality theorem, which further identifies the compact real Lie group G with the set of real points of a complex linear algebraic group $\bar{G}$. $\bar{G}$ is obtained by considering the algebra R of complex valued functions generated by the matrix coordinate functions of the representations of G: in contemporary terms, R is in a natural way a finitely-generated complex Hopf algebra, hence, the coordinate ring of an algebraic group $\bar{G}$. The main result is now that $\hat{\hat{G}}$ (the 'representations' of the set of representations of G, as above) is identified with $\bar{G}$.

At this point there are now really two kinds of dual objects to the compact Lie group G: the set $\hat{G}$ of representations of G, with its composition laws from tensor product and direct sum, and the algebra R of coordinate functions from representations, with its Hopf algebra structure. Both objects can be granted equal status if the group is assumed known (they both require knowing the category Mod(G)), but from the point of view we shall be concerned with here, the status of R as a dual object is a bit shaky. It appears that to describe R (it is, after all, an algebra of functions on G) one must know G, not just the category Mod(G). Appearances can be deceptive (we will see in Chapter 2 that R can be produced directly from Mod(G)), of course. Nonetheless, Chevalley's formulation of Tannaka's theorem via R (which came to be known as the algebra of representative functions on G) seems to have led to an emphasis on the algebraically more familiar object R and away from the rather awkward notion of 'representation' of $\hat{G}$.

This is, at least, what Hochschild & Mostow seem to tell us in their 1957 paper [7], the first of a series of ten papers [7, 8, 9, 10, 11, 12, 13, 14, 15, 16] published between 1957 and 1969 on the structure of the algebra $R(G)$ of representative functions on an analytic group G and its applications to the representation theory of G. There were other studies of Tannaka-type duality between 1941 and 1957, by Harish-Chandra in 1950 [5], and by Nakayama in 1951 [25], in various contexts. (Nakayama seems especially close to the major concern of this book when he says '. . . duality systems are defined purely in terms of representations . . . without appealing to the group and its group operation.') Nonetheless, it is certainly fair to say that the Hochschild–Mostow papers carry through the Chevalley formulation of Tannaka duality in terms of $R(G)$ to a complete theory that yields the structure of $R(G)$ for any analytic group G, explains why there is a duality in the compact case (Tannaka) and semisimple case (Harish-Chandra) as well as much more.

We make no attempt here to summarize the Hochschild–Mostow papers. Indeed, much of this book can be regarded as an exposition of those papers. There are, however, two points we shall note now: one is the discovery in [10] that if the analytic group G has a faithful representation then G has a structure of algebraic variety such that multiplications (on one side) by group elements are morphisms. The coordinate ring of such a structure sits inside $R(G)$ and yields a good description of $R(G)$ (see Chapter 4). The other point worth noting is that the last paper in the Hochschild–Mostow series [16] deals essentially with the question of determining G from Mod(G): given $R(G)$, as a Hopf algebra, how closely is G known?

Using $R(G)$ as the dual object to G is, as we noted above, somewhat disingenuous, if we want to regard G as unknown. A mechanism to reinsert Mod(G) (as a category only) into the theory appeared in Grothendieck's 1970 paper [4]. That paper is about detecting whether a homomorphism of discrete groups of finite type which induces an isomorphism on profinite completions is an isomorphism, not about analytic groups, but the main technical device used is to identify the profinite completion directly from the category of modules for the group, as the group of 'tensor automorphisms of the forgetful functor' (see (1.7) for a precise definition).

This idea was then developed in full generality in Saavedra-Rivano's 1972 lecture notes [27] on Tannakian categories: this is a category of vector spaces with tensor product, meeting a few technical conditions, which then turns out to be the category of all modules for its group of tensor automorphisms (natural equivalences of the forgetful functor to vector spaces preserving tensor product), provided that 'module' is taken in the appropriate sense. Here, that means regarding the group of tensor automorphisms as an affine group scheme so modules are group-scheme modules.

It is interesting to note that Tannaka's original ideas were quite close to Grothendieck's formulation. His dual object $\hat{G}$ consisted only of irreducible representations of G (this is permitted since the compactness of G implies that all representations are completely reducible) and a representation of $\hat{G}$ in his sense is a selection of a linear automorphism of each element of $\hat{G}$ compatible with tensor product and direct sum. Because of the irreducibility of the elements of $\hat{G}$ and Schur's lemma, this selection is automatically compatible with G-module homomorphisms, and extends to give exactly a tensor automorphism of Mod(G). The main point to keep in mind is that semisimplicity allows Tannaka to deal just with objects of Mod(G); the homomorphisms are essentially just given by scalars so are carried along for free.

Mod(G), when G is an analytic group, turns out to satisfy the axioms for a Tannakian category. So there is a group scheme such that Mod(G) is the category of modules for it. An affine group scheme is completely determined by its Hopf algebra of global functions, and it is not terribly surprising that the group-scheme for Mod(G) has $R(G)$ as its Hopf algebra. This was pointed out by Lubotzky in his 1978 PhD thesis from Bar-Ilan University.

It seems we are back to studying $R(G)$ again, but the difference is an important one: now $R(G)$ is constructed from Mod(G) without direct reference to G (we have even used the notation R(Mod(G)) at some points to emphasize the relation). And with the benefit of hindsight, we can see how to proceed: we start with Mod(G) and its group of tensor automorphisms, then construct $R(G)$ from that, and then go into the detailed study of $R(G)$. This in fact is the plan here, and we now turn to a detailed summary of the book.

3

This chapter-by-chapter summary is intended to let the potential reader know what is in store in a general way, as well as to let him know what the author considers to be the dominant themes of the work and how they are developed. Readers in search of specific results should consult the summaries which conclude each chapter.

Chapter 1 is entitled ‘Definitions and Examples’. The definitions are those of representation and module for an analytic group G, and for $\mathrm{Mod}(G)$, its group of tensor automorphisms $\mathrm{Aut}_{\otimes}(\mathrm{Mod}(G))$, and a canonical group homomorphism $\sigma: G \to \mathrm{Aut}_{\otimes}(\mathrm{Mod}(G))$. The examples (there are five labeled A to E) calculate $\mathrm{Mod}(G)$, and as far as possible $\mathrm{Aut}_{\otimes}(\mathrm{Mod}(G))$, from first principles, for $G = \mathrm{GL}_1\mathbb{C}$ (multiplicative group of complex numbers), $G = \mathbb{C}$ (additive group of complex numbers), $G = \mathbb{C} \rtimes \mathbb{C}^*$ (semidirect product of multiplicative group of complexes $\mathbb{C}^*$ acting on $\mathbb{C}$), $G = \mathbb{C} \rtimes \mathbb{C}$ (semidirect product of $\mathbb{C}$ acting on $\mathbb{C}$ via exponential $\mathbb{C} \to \mathrm{GL}_1\mathbb{C}$), and $G = \mathbb{C}^{(2)} \rtimes \mathbb{C}$ (semidirect product of $\mathbb{C}$ acting on $\mathbb{C}^{(2)}$ via

$$\mathbb{C} \to \mathrm{GL}_2\mathbb{C},\ t \mapsto \begin{bmatrix} e^{\alpha t} & 0 \\ 0 & e^{\beta t} \end{bmatrix},$$

$\alpha, \beta \in \mathbb{C}$).

These groups are all elementary in the sense that they are at worst two-stage solvable. Determining their modules then is quickly reduced to finding modules for $\mathbb{C}^*$ or $\mathbb{C}$, which are the first two examples, subject to some compatibility rules. Now $\mathbb{C}^*$-modules are the same as graded vector spaces, and $\mathbb{C}$-modules are the same as vector spaces with a designated endomorphism, so finding the modules can be accomplished. Since the tensor product structure of $\mathrm{Mod}(G)$ is central to our discussions, we also search for what might be called generators of $\mathrm{Mod}(G)$ as a tensored category, and find them. Tensor automorphisms are determined by their values on these generators, and this gives us a handle on $\mathrm{Aut}_{\otimes}(\mathrm{Mod}(G))$.

The examples are elementary, but they point out most of the possibilities that occur in general. Here are the answers: (A) for $G = \mathrm{GL}_1\mathbb{C}$, σ is an isomorphism; (B) for $G = \mathbb{C}$, σ is one–one and $\mathrm{Aut}_{\otimes}(\mathrm{Mod}(G)) = \sigma(G) \times \mathrm{Hom}_{\mathrm{grp}}(\mathbb{C}, \mathbb{C}^*)$ (the second factor is the abstract group homomorphisms from $\mathbb{C}$ to $\mathbb{C}^*$; (C) for $G = \mathbb{C} \rtimes \mathbb{C}^*$, σ is again an isomorphism; (D) for $G = \mathbb{C} \rtimes \mathbb{C}$, σ is one–one and $\mathrm{Aut}_{\otimes}(\mathrm{Mod}(G)) = \sigma(G) \rtimes \mathrm{Hom}_{\mathrm{grp}}(\mathbb{C}, \mathbb{C}^*)$. (Example E is discussed below separately.)

Notice that it is for the vector group $\mathbb{C}$ (Example B) where the group of tensor automorphisms exceeds G, and the excess, it turns out, comes from the additive character group $\mathrm{Hom}(G, \mathbb{C})$ of analytic homomorphisms from G to $\mathbb{C}$; $\mathrm{Hom}(G, \mathbb{C})$ is the '$\mathbb{C}$' of the second factor of $\mathrm{Aut}_{\otimes}$. Example C has a large vector subgroup, but $\mathrm{Hom}(G, \mathbb{C}) = 1$ here and σ is an isomorphism. In Example D again, $\mathrm{Aut}_{\otimes}$ exceeds G, and the excess is again measured by the additive character group. In Examples A, B, and C, G can also be regarded as an algebraic group and these are the cases where $\sigma(G)$ is a direct factor of $\mathrm{Aut}_{\otimes}(\mathrm{Mod}(G))$. In general, one sees later that the excess (if any) of $\mathrm{Aut}_{\otimes}$ over $\sigma(G)$ is always measured by additive characters, and that of $\sigma(G)$ is a direct factor in $\mathrm{Aut}_{\otimes}$ exactly when G is algebraic.

The fifth example (actually it is a two parameter family of examples) is used somewhat differently. If the pair of parameters are algebraically the same in two of the groups (in the sense that they span isomorphic rational vector spaces), then a functor can be exhibited showing that the categories of modules for the two groups are equivalent. The parameters must be much closer for the groups to be isomorphic, so we produce lets of examples of nonisomorphic groups with the same module categories. This warns us, of course, to focus on $\mathrm{Mod}(G)$, not G, in our constructions later.

Chapter 2, 'Representative Functions', deals with the formal description of the category $\mathrm{Mod}(G)$ of finite-dimensional modules for the analytic group G. The representative functions on G coming from the module V are the functions on G, $g \mapsto f(gv)$ (for fixed $v \in V$ and $f \in V^*$), and $R(G)$ denotes the set of all such as f, v, and V vary over $\mathrm{Mod}(G)$. Now $R(G)$ also admits an intrinsic description as the set of holomorphic functions on G whose G-

translates span a finite-dimensional vector space. Under pointwise addition and multiplication of functions $R(G)$ is a complex algebra, in fact an integral domain.

It has further structure: using the group operators of G, $R(G)$ becomes a Hopf algebra. For example, if $f \in R(G)$ then the comultiplication sends f to $\sum h_i \otimes k_i$ if $f(xy) = \sum h_i(x)k_i(y)$ for all x and y in G. Both the algebra and coalgebra structure on $R(G)$ are defined using G. (This will be remedied later.) Moreover, there is a natural equivalence between the category Mod(G) and the category of finite-dimensional comodules for $R(G)$. Thus, knowledge of $R(G)$, as a Hopf algebra, determines Mod(G).

The converse of this assertion also holds: the Hopf algebra $R(G)$ can be produced directly from the category Mod(G). This requires describing $R(G)$ other than as an algebra of functions on G, of course. It turns out that this can be done by describing $R(G)$ as an algebra of functions on $\mathrm{Aut}_\otimes(\mathrm{Mod}(G))$. This is accomplished via several intermediate steps.

Associated to the Hopf algebra $R(G)$ are two other groups, $\mathcal{G}(G)$, which is the group of $\mathbb{C}$-algebra homomorphisms from $R(G)$ to $\mathbb{C}$, and Propaut($R(G)$), which is the group of $\mathbb{C}$-algebra automorphisms of $R(G)$ commuting with right G-translations. Both groups are isomorphic. Moreover, there are homomorphisms $\mathrm{Aut}_\otimes(\mathrm{Mod}(G)) \to \mathrm{Propaut}(R(G))$ and $\mathcal{G}(G) \to \mathrm{Aut}_\otimes(\mathrm{Mod}(G))$ which are compatible with this isomorphism, so that all three groups are isomorphic. Finally, one shows that every G-module is naturally an $\mathrm{Aut}_\otimes$-module and that $R(G)$ is the algebra of all functions on $\mathrm{Aut}_\otimes(\mathrm{Mod}(G))$ of the form $a \mapsto f(av)$ where $v \in V$ and $f \in V^*$ for some G-module V. Thus, we conclude that $R(G)$ can be produced, as a Hopf algebra, from Mod(G). Moreover, we end up with a handle on the structure of $\mathrm{Aut}_\otimes(\mathrm{Mod}(G))$: we can think of it as the group $\mathcal{G}(G)$, so to determine $\mathrm{Aut}_\otimes$ we can confine our attention to the Hopf algebra $R(G)$, which is a single (large) object on which G acts, rather than have to deal with the entire category Mod(G). Now $\mathcal{G}(G)$ also receives a map from G, and we can regard $G \to \mathcal{G}(G)$ as a kind of 'universal representation' of G, in a sense which will be made clear below. Once we start dealing with this map, to be sure, we have moved our point of view in two senses: first, we are dealing directly with G again – this is not a map determined from the category Mod(G) – and second, we have shifted from modules to representations. In the module point of view the categorical properties of the set of all representations became tractible, but the group structure of G is in the background. Now, in the representation point of view, it comes to the front.

If $\rho : G \to \mathrm{GL}(V)$ is a representation and $\bar{G}$ is the Zariski-closure of $\rho(G)$ in the algebraic group $\mathrm{GL}(V)$, then $\rho(G)$ becomes a Zariski-dense analytic subgroup of the algebraic group $\bar{G}$. This is the situation studied in Chapter 3, 'Analytic subgroups of algebraic groups'. The main point of the chapter is to see that such subgroups are pretty close to being algebraic themselves.

An analytic subgroup of an algebraic group is an analytic group with a faithful representation. It turns out that the analytic groups G with a faithful representation are the groups which have a nucleus: that is, a connected, simply-connected solvable normal subgroup K such that G/K is reductive (reductive means faithfully representable and all modules semisimple). This is actually one of the last results of the chapter; it begins by considering groups with nuclei. If G is one such, and K is a nucleus, then there is a reductive subgroup P of G with $G = KP$ and $K \cap P = \{e\}$. Next, it is shown that if G has such a decomposition $G = KP$ then G can be embedded in an algebraic group $\bar{G}$ via an analytic embedding f such that $f(G)$ is Zariski-dense in $\bar{G}$, and there is a torus T in $\bar{G}$ such that $\bar{G} = f(G)T$, $T \cap f(G) = e$, and T centralizes $f(P)$. We call the triple $(\bar{G}, f, T)$ a split hull of G.

We pause a moment to reflect on the importance of the torus T. Suppose we have embedded G in $\mathrm{GL}(V)$ and chosen a basis so that the solvable subgroup K is in upper triangular form. The intersection of K with the diagonal maximal torus of $\mathrm{GL}(V)$ is an analytic, but possibly not algebraic, subgroup and in taking its Zariski-closure T appears. If we look at the diagonal entries as functions on K, they have the form $x \mapsto \mathrm{e}^{a(x)}$ where $a : K \to \mathbb{C}$ is an analytic homomorphism. It is the presence of such functions that separates the analytic theory from the algebraic theory, and the torus T is a group-theoretical record that there are such analytic but not algebraic characters present.