Arch. Hist. Exact Sci. (2012) 66:295–358
DOI 10.1007/s00407-012-0095-8

The early proofs of the theorem of Campbell, Baker, Hausdorff, and Dynkin

Rüdiger Achilles · Andrea Bonfiglioli

Received: 30 December 2011 / Published online: 3 April 2012

Abstract The aim of this paper is to provide a comprehensive exposition of the early contributions to the so-called Campbell, Baker, Hausdorff, Dynkin Theorem during the years 1890–1950. Related works by Schur, Poincaré, Pascal, Campbell, Baker, Hausdorff, and Dynkin will be investigated and compared. For a full recovery of the original sources, many mathematical details will also be furnished. In particular, we rediscover and comment on a series of five notable papers by Pascal (Lomb Ist Rend, 1901–1902), which nowadays are almost forgotten.

1 Introduction

The algebraic result named after (some or all of) the mathematicians Henry Frederick Baker, John Edward Campbell, Eugene Borisovich Dynkin, and Felix Hausdorff is one of the most versatile results originating from the early theory of groups of transformations. This result states that, in the algebra of the formal power series in two non-commuting indeterminates x and y, the series naturally associated to $\log(e^x e^y)$ is a series of Lie polynomials in x and y; for instance, the first few terms are

$$\log(e^x e^y) = x + y + \tfrac{1}{2}[x, y] + \tfrac{1}{12}[x, [x, y]] - \tfrac{1}{12}[y, [x, y]] + \cdots$$

Communicated by: Umberto Bottazzini.

R. Achilles (✉) · A. Bonfiglioli
Dipartimento di Matematica, Università di Bologna, Piazza di Porta S. Donato 5, 40126 Bologna, Italy
e-mail: rudiger.achilles@unibo.it

A. Bonfiglioli
e-mail: andrea.bonfiglioli6@unibo.it

During the whole twentieth century, this result (which, for convenience, we henceforth refer to as "CBHD," after its major contributors, or generically as "the Exponential Theorem") has yielded the following applications:

- to *Physics*, especially with the consolidation of Quantum and Statistical Mechanics—see the early applications from the 1960s and 1970s: Białynicki-Birula et al. (1969), Dragt and Finn (1976), Eriksen (1968), Friedrichs (1953), Gilmore (1974), Kumar (1965), Mielnik and Plebański (1970), Murray (1962), Suzuki (1977), Wei (1963), Weiss and Maradudin (1962), Wichmann (1961), and Wilcox (1967); see also the numerous recent papers in Physics journals related to CBHD: Blanes et al. (1998), Blanes et al. (2009), Bose (1989), Casas (2007), Casas and Murua (2009), Klarsfeld and Oteo (1989a), Klarsfeld and Oteo (1989b), Kobayashi et al. (1998), Kolsrud (1993), Moan (1988), Moan and Noteo (2001), Oteo (1991), and Reinsch (2000); and finally see the investigations concerning the relevant convergence and optimization problems: Blanes and Casas (2004), Day et al. (1991), Moan and Niesen (2008), Newman et al. (1989), Richtmyer and Greenspan (1965), and Thompson (1982, 1989);
- to *Group Theory*: Magnus (1950), Magnus et al. (1966), and Michel (1976);
- to the *Analysis of Linear PDEs*: See the seminal works from the 1970s and 1980s Folland (1975), Folland and Stein (1982), Hörmander (1967), Rothschild and Stein (1976), and Nagel et al. (1985);
- to the structure theory of *Lie Groups and Lie Algebras*: Abbaspour and Moskowitz (2007), Bourbaki (1972), Duistermaat and Kolk (2000), Godement (1982), Gorbatsevich et al. (1997), Hall (2003), Hausner and Schwartz (1968), Hilgert and Neeb (1991), Hochschild (1965), Hofmann and Morris (2006), Jacobson (1962), Rossmann (2002), Sagle and Walde (1973), Sepanski (2007), Serre (1965), and Varadarajan (1984);
- more recently, to *Numerical Analysis* (in geometric integration): Hairer et al. (2006), Iserles et al. (2000), Iserles and Nørsett (1999), and McLachlan and Quispel (2002).

Furthermore, a special comment has to be made on the role played by the CBHD Theorem also within the context of *infinite dimensional* Lie groups, nowadays intensively studied; see Neeb (2006) for a detailed survey. For example, in this context, the so-called Baker–Campbell–Hausdorff groups are particularly significant: for related topics see Birkhoff (1936, 1938), Boseck et al. (1981), Czyż (1989, 1994), Dynkin (1950), Glöckner (2002a,b,c), Glöckner and Neeb (2003), Gordina (2005), Hilgert and Hofmann (1986), Hofmann (1972, 1975), Hofmann and Morris (2007), Hofmann and Neeb (2009), Neeb (2006), Omori (1997), Robart (1997, 2004), Schmid (2010), Van Est and Korthagen (1964), Vasilescu (1972), and Wojtyński (1998).

Finally, the interest in providing new and simpler proofs of the CBHD Theorem seems to have been renewed throughout the century: see e.g., Yosida (1937), Cartier (1956), Eichler (1968), Djoković (1975), Veldkamp (1980), and Tu (2004).

As attention focussed more and more on the applications, and after the CBHD Theorem had become the well-consolidated result of non-commutative Algebra, critique of the early contributions lost its appeal. But this resulted, from the very beginning, in a very confused and uncertain attribution of the "fatherhood" of the theorem, which was named sometimes after Campbell, Baker, and Hausdorff, some other times after

Campbell and Hausdorff, or also after Hausdorff solely. As we shall discuss in detail, the last two choices are very questionable and seem dictated mainly by tradition.

Needless to say, the contributions given by the forerunners of the theorem, Ernesto Pascal, Jules Henri Poincaré, and Friedrich Heinrich Schur, have to a great extent been forgotten. This is definitely true of Pascal's and of Schur's contribution, whereas that by Poincaré luckily has been brought to light by Schmid (1982) after many years of oblivion (see also Ton-That and Tran 1999). We point out that the paper (Schmid 1982) also contains some salient points on the whole history of the Exponential Theorem, of which we took advantage in writing this paper. Yet, a full recovery of the prior results (of Pascal, Poincaré, and Schur) and a comparison of the actual contributions by the authors, whose names have been usually attached to the theorem (Baker, Campbell, Hausdorff, and Dynkin) have not been provided in the literature.[1]

The aim of the present paper is to provide these missing investigations and comparisons of the early contributions to the Exponential Theorem furnished by all the cited authors. In particular, we rediscover a long list of works by the Italian mathematician Pascal (1901a,b, 1902a,b,c, 1903), which had been of decisive importance for the history of this theorem.

Since the original sources are not all easily accessible,[2] and since not only the names of some contributors but also the contributions themselves have been lost, we discuss the mathematical details. In doing this, we hope that our article will be helpful to those who wish to rediscover the contents of the original papers (these papers being written in many different languages: English, French, German, Italian, and Russian). The paper is indeed intended for the specialists in the fields of applications mentioned in the *incipit*, who may find use for these mathematical contents, as well as for experts in the History of Mathematics. Even if this paper is based on the works of the early contributors to the Exponential Theorem, the analysis and comparison of these contributions are carried out here for the first time. Moreover, our intention is to ease the access to these papers by explaining their contents in modern language.

The paper is organized as follows. Section 2 recalls the roots of the CBHD Theorem, as a result originating implicitly in Sophus Lie's theory of continuous groups of transformations. Subsequently, we analyze and compare the works of Schur, Campbell, Poincaré, Pascal, Baker, Hausdorff, and Dynkin on the Exponential Theorem: this is done from Sect. 3 to Sect. 9. Section 10 records the commentaries by Bourbaki (1972) and by Hausdorff (1906) on the early contributions. As we shall see, the authoritative influence of Bourbaki and Hausdorff affected to a great extent the subsequent forgetting of the other authors' contributions.

Without being exhaustive in this Introduction, we furnish a *précis* of some (less-known or maybe unknown) results presented in this paper.

[1] For example, Dieudonné (1982, Chapter BII, p. 189) quotes the names of Campbell, Dynkin, Hausdorff, Poincaré, and Schur (omitting Baker and Pascal) in a list of those who "contributed substantially" to the theory of Lie's groups. Curiously, the historically interested mathematicians, Borel (2001, Chapter I, p. 6) and Dieudonné (1974, Chapter XIX, p. 222), refer to the Exponential Theorem as the "Campbell–Hausdorff formula," still omitting any reference to Baker.

[2] This is particularly true for the Italian papers by Pascal (1901a,b, 1902a,b,c).

(i) As Engel reveals to us in the review[3] of Pascal (1901b), Sophus Lie himself embarked on the search for an explicit formula providing, for any pair of infinitesimal transformations X_1, X_2, the infinitesimal transformation X_3 such that $e^{X_3} = e^{X_2} \circ e^{X_1}$, but, as Engel says, Lie abandoned the undertaking since he saw no way of proving the *convergence* of the series representing X_3 in terms of X_1, X_2. It will be the contribution of Hausdorff (1906) to solve the convergence problem in a convincing way.

(ii) In proving the Third Fundamental Theorem of Lie by a direct approach, Schur (1890a,b) constructed explicit infinitesimal transformations with prescribed structure constants c_{ijk}, with the use of some converging multivariate power series depending only on the c_{ijk} and the Bernoulli numbers.[4] Quoting Schmid (1982, p. 177), this is a version of the "Campbell–Hausdorff formula in disguise." In fact, a few years later, Pascal (1902c) will show how Schur's explicit formulas can be obtained by means of the Exponential Theorem, thus unveiling in an incontrovertible way the deep link between Schur's formulas and the future CBHD Theorem.

(iii) As will be seen in due course, the summands in the expansion of $\log(e^x e^y)$ containing x precisely once are given by

$$z_1(x,y) := \sum_{n=0}^{\infty} \frac{\mathbf{B}_n}{n!} (\mathrm{Ad}\, y)^n(x) =: \frac{\mathrm{Ad}\, y}{e^{\mathrm{Ad}\, y} - 1}(x), \tag{1}$$

where the $\mathbf{B}_n$ are the Bernoulli numbers. This series will play central roles in the proofs by Campbell, Pascal, Baker, and Hausdorff. The first one to introduce z_1 was Campbell (1897a), who explicitly employed this series to construct, recursively, a sequence approaching $\log(e^x e^y)$. Campbell's proof of the convergence of this sequence unfortunately leaves much unsaid, and his successors will acknowledge him only for having posed for the first time the *problem* of the existence of a universal Lie-polynomial series z in x, y such that $e^z = e^x e^y$: the so-called *Campbell's Problem*. Campbell gives *in nuce* a contribution to the later "ODE approach" to the proof of the Exponential Theorem, since he first obtains the identity:

$$e^{tx} e^y = e^{y + t\, z_1(x,y)} + \mathcal{O}(t^2), \quad \text{as } t \to 0.$$

[3] Here and in the sequel, when we cite reviews, it is understood that we refer to "Jahrbuch über die Fortschritte der Mathematik," at http://www.emis.de/projects/JFM.

[4] Each author uses his own notation and definition for the Bernoulli numbers, which we shall respect to preserve the flavor of the original formulas. For clarity, we introduce our separate conventional notation for the numbers $\mathbf{B}_n$ (which we call *Bernoulli numbers*) defined unambiguously by the following generating function:

$$\frac{z}{e^z - 1} = \sum_{n=0}^{\infty} \frac{\mathbf{B}_n}{n!} z^n \qquad (z \in \mathbb{C} : \ |z| < 2\pi).$$

We shall explicitly write the relationship occurring between these $\mathbf{B}_n$ and the constants used by the other authors.

(iv) The first author who really exploited the ODE technique in an efficient way was Poincaré (1900), who characterized the solution $z(t)$ of $e^{z(t)} = e^x e^{ty}$ by the (formal) ordinary differential equation:

$$z'(t) = \frac{\operatorname{Ad} z(t)}{1 - e^{-\operatorname{Ad} z(t)}}(y), \quad z(0) = x.$$

Note the hidden presence of the Bernoulli numbers, for

$$\frac{z}{1 - e^{-z}} = 1 + \frac{z}{2} + \sum_{n=1}^{\infty} \frac{\mathbf{B}_{2n}}{(2n)!} z^{2n}.$$

In the setting of a Lie group of transformations, this ODE has a genuine meaning and is solved by Poincaré via the residue calculus, in such a way that the Lie-series nature of $z(1) = \log(e^x e^y)$ is exhibited. What Poincaré actually failed to point out was the existence of a *universal* Lie-series expansion for $z(1)$ in terms of x and y, independent of any group context. Hausdorff and Bourbaki will later criticize Poincaré for his repeated reference to the group setting instead of using a symbolic approach. Notwithstanding this, in solving what he calls "le Problème de Campbell," Poincaré forged nothing less than *the universal enveloping algebra* of a Lie algebra, and at least *implicitly* in his technique, we can find a constructive way to turn $\log(e^x e^y)$ into a Lie series. This is done in Poincaré (1900) by a subtle "symmetrization process" of polynomials which we shall analyze more in detail (see also Schmid 1982 and Ton-That and Tran 1999).

(v) It was Pascal (1901a,b) who carried forward this symmetrization process, likely unaware that he was handling some computations that Poincaré properly inscribed a year before in the setting of enveloping algebras. The similarity between the approaches of Poincaré and Pascal in attacking the Exponential Theorem is so striking that we could name this approach *the Poincaré–Pascal technique*, which, for its originality, would deserve to be rediscovered and formalized. Roughly speaking, Pascal de-constructs $e^x e^y$ into $\sum_{m,n\geq 0} \frac{x^m y^n}{m!\,n!}$ and then he re-constructs this double series (via a certain symmetrization of polynomials and a delicate handling of constants) into $\sum_{h\geq 0} \frac{z^h}{h!}$, for a suitable formal series z for which he also describes an iterative formula in terms of Lie brackets of x, y and the Bernoulli numbers. Conscious of having thus provided a nodal contribution to the understanding of the composition of two exponentials in transformation group theory, Pascal deferred the convergence problem. For this omission, and for the massive computations employed, Engel (in reviewing Pascal 1901b, 1902c) and Hausdorff later criticized Pascal, while Bourbaki mentioned Pascal's name in passing only. Regardless of this legitimate criticism, Pascal's contribution to the CBHD Theorem is, without doubt, fundamental, and the present paper is an occasion to popularize it after more than one hundred years of oblivion.

(vi) The main step toward a symbolic proof of the Exponential Theorem (this is the very name used by Baker 1901, 1905, §3, p. 35, see also *Exponentialformel*, in the title of Hausdorff 1906) comes with Baker's and Hausdorff's proofs. Some core computations and the main formulas obtained in the two papers Baker (1905) and Hausdorff (1906) are so similar that it is not easy to realize that they are independent of each other. Ipso facto, Hausdorff's study (1906) was published one year after (Baker 1905), but no mention of Baker (1905) is made in Hausdorff (1906), though the analogies are several. Campbell's series $z_1(x, y)$ in (1) reappears in a very "quantitative" way, and Baker and Hausdorff prove that it generates $\log(e^x e^y)$ in the following form:

$$e^x e^y = e^z, \quad \text{where } z = \sum_{n=0}^{\infty} \frac{1}{n!} \left(z_1(x, y) \frac{\partial}{\partial y} \right)^n y. \tag{2}$$

Here $z_1 \frac{\partial}{\partial y}$ is a sort of "substitutional operation" (as Baker called it, later also named *polar derivative*), more precisely it is the derivation of the associative algebra of the formal power series in x, y which is defined by $z_1 \frac{\partial}{\partial y}(x) = 0$ and $z_1 \frac{\partial}{\partial y}(y) = z_1$. Owing to the puzzling notations used by Baker in proving (2), when compared to the undoubted clarity of Hausdorff's exposition, Bourbaki will consecrate Hausdorff as the only reliable source. Despite this fact, the appropriate naming for (2) seems to us to be *the Baker–Hausdorff Formula*. Hence, the designation "Campbell–Hausdorff Formula" (in use, in books at least, seemingly after Jacobson 1962) is more an abused tradition (so entrenched that one can do nothing but get along with it) than a reference to the actual history of the theorem. Furthermore, if one looks for something that may be called a "Formula," Hausdorff provided at least another two of them, whereas Campbell provided none (so what is the "Campbell–Baker–Hausdorff Formula" after all?). Among Hausdorff's contributions, we include the successful recursion formula which can be obtained by the identity:

$$\left(x \frac{\partial}{\partial x} \right) z + \left(y \frac{\partial}{\partial y} \right) z = [x, z] + \frac{\operatorname{Ad} z}{1 - e^{-\operatorname{Ad} z}}(x + y),$$

and by decomposing z into homogenous summands with respect to x and y jointly. This recursion formula will return in Djoković (1975) and in Varadarajan (1984).

(vii) Rather than speaking of a Formula of Campbell, Baker, and Hausdorff, the "greatest common divisor" shared by the contributions of these three mathematicians is the following statement (more a qualitative result than an actual formula): *In the associative algebra (over a field of characteristic zero) of the formal power series in two non-commuting indeterminates x and y, the series related to* $\log(e^x e^y)$ *is a series of Lie polynomials. This may be referred to as the Campbell–Baker–Hausdorff Theorem.*

(viii) Proceeding with the above (provocative) disquisition about "formulas," probably the most important one is due to someone who is not even mentioned in the naming "Campbell–Baker–Hausdorff Formula." It is indeed Dynkin (1947) who (for the very first time and more than 40 years after Hausdorff's paper) furnished *in closed form* the sought-for explicit presentation of $\log(e^x e^y)$ in terms of iterated

Table 1 Comprehensive references for the early proofs of the CBHD Theorem (range of period 1890–1950)

Year	Paper
1890	Schur (1890a,b)
1891	Schur (1891)
1893	Schur (1893)
1897	Campbell (1897a,b)
1898	Campbell (1898)
1899	Poincaré (1899)
1900	Poincaré (1900)
1901	Baker (1901)
1901	Pascal (1901a,b)
1901	Poincaré (1901)
1902	Baker (1902)
1902	Pascal (1902a,b,c)
1903	Baker (1903)
1903	Campbell (1903)
1903	Pascal (1903)
1905	Baker (1905)
1906	Hausdorff (1906)
1937	Yosida (1937)
1947	Dynkin (1947)
1949	Dynkin (1949)
1950	Dynkin (1950)

brackets, the so-called *Dynkin's Formula*. Nowadays, it is customary to mention Dynkin's name within the contributions to the Exponential Theorem only when his representation is involved. In fact, Dynkin's contribution is paramount for other distinguished reasons which we shall discuss in detail in Sect. 9 (and which have inspired our acronym CBHD); among these reasons is the merit of having provided another proof of the Exponential Theorem (completely different from the preceding ones, see Dynkin 1949) enlightening all the *combinatorial* aspects behind the theorem, together with a proof of the convergence matter, far more natural and simpler than Hausdorff's.

(ix) After 30 years after the mentioned paper (Hausdorff 1906), Yosida contributed to clarify some passages in Hausdorff's main argument, dropping the symbolic polar derivatives in favor of the usual differentiation with respect to real (or complex) parameters.

To ease the reading, in Table 1 we furnished a chronological list of the original contributions we shall be dealing with (in the range of period 1890–1950).

2 The origin of the problem

The main theorem we are concerned with in this paper has its roots in Sophus Lie's theory of the so-called finite continuous groups of transformations. As is well known, what we now call "Lie groups" are quite different from what Lie himself occasionally referred to as "his groups" and what he studied during the second half of the nineteenth century. Since we do not need Lie's original theory elsewhere in the paper, we confine ourselves in recalling a (simplified) definition of continuous group of transformations (for more details, see Hawkins 1989, 1991, 2000; see also Montgomery and Zippin 1955 or Varadarajan 1984, §2.16).[5]

To this end, let E^m denote (real or complex) m-dimensional Euclidean space and let $A \subseteq E^r$ be an open neighborhood of 0, and $\Omega \subseteq E^n$ be some domain. Then a *finite continuous group of transformations* is a family of maps indexed by $a \in A$

$$f(\cdot, a) : \Omega \to \Omega, \qquad x \mapsto x' = f(x, a),$$

such that the following properties are satisfied (together with others omitted for brevity; 0 is chosen as identity element for simplicity):

(i) the maps $(x, a) \mapsto f(x, a)$ are analytic and $f(x, 0) = x$ for every x;
(ii) for every couple $(a, b) \in A \times A$ there exists *a unique* $\varphi(a, b) \in A$ such that

$$f(f(x, a), b) = f(x, \varphi(a, b)), \quad \text{for every } x \in \Omega. \tag{3}$$

It is not difficult to see that the map $(a, b) \mapsto \varphi(a, b)$ then defines a (local) analytic group, so that, in modern words, a transformation group is nothing but the germ of a group action on a domain of $\mathbb{R}^n$ or $\mathbb{C}^n$.

Throughout the paper, when we speak of a group of transformation (or of a transformation group), we always refer to the above definition.

A key role is played by the so-called infinitesimal transformations of the transformation group, something very similar to what we would now call the relevant Lie algebra. For every $j = 1, \dots, r$, every smooth function F on Ω and every $x \in \Omega$, let us set

$$X_j F(x) = \sum_{i=1}^{n} \frac{\partial f_i(x, 0)}{\partial a_j} \cdot \frac{\partial F(x)}{\partial x_i}. \tag{4}$$

The action on F of the differential operator $\sum_{j=1}^{r} a_j X_j$ provides, as is easily seen, the limit

$$\lim_{\varepsilon \to 0} \frac{F(f(x, \varepsilon a)) - F(x)}{\varepsilon};$$

thus it furnishes a sort of "weight" of the infinitesimal transformation in the "direction" of the parameter $a = (a_1, \dots, a_r)$.

[5] It is also of interest to refer to some books on ODEs from the first decades of the twentieth century, to see how groups of transformations (originally conceived by Lie for differential equation purposes) still played a role in the ODE theory of those years: see, e.g., Campbell (1903), Cohen (1931), Ince (1927), and Pascal (1903). A beautiful exposition of the ODE/PDE contents of the theory of transformation groups is contained in Eisenhart (1933).

The Second Fundamental Theorem of Lie ensures that the span of the $X_1, \dots, X_r$ is closed under the usual bracket operation, i.e., there exist constants c_{ijk} (called the *structure constants* of the transformation group) such that

$$[X_i, X_j] = \sum_{k=1}^{r} c_{ijk} X_k, \quad \text{for every } i, j = 1, \dots, r. \tag{5}$$

The converse of this fact (part of the Second Fundamental Theorem too) is less trivial, stating that, if the independent analytic vector fields $X_1, \dots, X_r$ on the domain $\Omega \subseteq \mathbb{R}^n$ span a finite-dimensional Lie algebra, then they are the infinitesimal transformations of some group of transformations. The arguments to prove this fact make use of the one parameter groups generated by the family of vector fields $X_1, \dots, X_r$, namely, the solutions to the ODE system

$$\begin{cases} \dot{\gamma}(t) = A(\gamma(t)) \\ \gamma(0) = x \end{cases} \quad \text{where} \quad A = a \cdot X := a_1 X_1 + \dots + a_r X_r, \tag{6}$$

$a_1, \dots, a_r$ being scalars and $x \in \Omega$. If the parameters a_i are small enough, the position

$$f(x, a) := \gamma(1) \tag{7}$$

makes sense, and it defines the desired transformation group. The proof of the existence of the relevant group law $\varphi(a, b)$ is not so direct: φ is the solution of suitable (first order) PDE problems integrability of which derives from the existence of constants c_{ijk} satisfying (5).[6]

Despite the concealed nature of φ, *the above facts contain the core of the CBHD Theorem* as we now describe. Indeed, the Maclaurin expansion of γ is

$$\gamma(t) \sim \sum_{k=0}^{\infty} \frac{t^k A^k(x)}{k!} =: e^{tA}(x), \tag{8}$$

justifying the exponential notation for $\gamma(t)$. From (7) and from the group property (3), it follows that the composition of two exponentials must obey the following law (here $\circ$ is the composition of functions, see also the notation in (6)):

$$e^A \circ e^B = e^C, \quad \text{where } A = a \cdot X, B = b \cdot X, C = \varphi(a, b) \cdot X. \tag{9}$$

Baker (1905, p. 24) expresses very clearly a notable intuition about the product of two exponentials: "*it is an obvious suggestion of Lie's theory that the product $e^A e^B$ is of the form e^C, where C is a series of alternants* [i.e., of brackets] *of A and B*," even under the only assumption that A and B are any "*two non-commutative quantities.*"

6 For a more exhaustive exposition of this approach to the Second Fundamental Theorem, having an historical interest and tracing back to Lie himself, see, e.g., Schur (1890a), Pascal (1903), and Eisenhart (1933).

In facing this question, Baker was preceded by Campbell (1898, see the *incipit*), who, for the first time, proposed to investigate the quantity $e^A e^B$ from a purely symbolic point of view. Later, this was to be called *Campbell's Problem*, see Poincaré (1900, p. 232) and Hausdorff (1906, p. 20).

It is then clear that the study of the "product" of two exponentials, originating from Lie's theory of transformation groups, deserved independent attention to be understood in its generality. Lie did not pay much attention to this problem, for the equation $e^A e^B = e^C$ was, roughly speaking, contained in the axiomatic identity (3) of his theory (or, more precisely, in the proof of his Second Fundamental Theorem). What Lie did not prove, however, was the fact that the law of formation of C as a series of brackets of A, B is in fact "universal":

$$C = A + B + \tfrac{1}{2}[A, B] + \tfrac{1}{12}[A[A, B]] - \tfrac{1}{12}[B[A, B]] + \tfrac{1}{24}[A[B[B, A]]] + \cdots,$$

and that the structure constants c_{ijk} of the given group intervene only if one wants to write C in the basis $\{X_1, \dots, X_r\}$ (this basis decomposition being immaterial for the study of the "symbolic" identity $e^A e^B = e^C$). The universal symbolic nature of C is the topic that Campbell, Baker, Hausdorff, and Dynkin eventually addressed, thus untying the Exponential Theorem from the original group context where it was born, and lending it an interest in its own right.

Let us finally recall Lie's Third Fundamental Theorem, originally stated as follows: **Lie's Third Theorem**. *Given r^3 real constants $\{c_{ijk}\}_{i,j,k\le r}$ satisfying the conditions*

$$c_{ijk} = -c_{jik}, \qquad \sum_{s=1}^{r} (c_{ijs}c_{skh} + c_{jks}c_{sih} + c_{kis}c_{sjh}) = 0, \tag{10}$$

for every $i, j, h, k = 1, \dots, r$, there exists a transformation group whose structure constants are the c_{ijk}, i.e., such that the infinitesimal transformations $X_1, \dots, X_r$ of this group satisfy

$$[X_i, X_j] = \sum_{k=1}^{r} c_{ijk} X_k. \tag{11}$$

Throughout the sequel, when we speak of Lie's Third (Fundamental) Theorem, we tacitly make reference to the above result, which is significantly different from what is nowadays understood as the Third Fundamental Theorem of Lie. (Roughly speaking, the original result is a *local* version of the modern one.[7])

[7] The modern, *global* version of Lie's Third Theorem is the following one: *For every finite dimensional real Lie algebra $\mathfrak{g}$ there exists a Lie group G such that $\mathfrak{g}$ is isomorphic to the Lie algebra of G.* Here the notion of Lie group is, obviously, the modern one. The proof of this global form of the theorem is of a different caliber, much more involved than the proof for the original version: some arguments require, e.g., Ado's Theorem asserting that every finite dimensional Lie algebra over a field of characteristic zero has a faithful linear representation. For a very recent functorial version of Lie's Third Theorem, see Hofmann and Morris (2005).

Besides precursory studies by Schur, the link between Lie's Third Theorem and the Exponential Theorem was first fully understood (as we shall explain) by Poincaré, whose proof of the Third Theorem is more or less the modern one we would give today starting from the CBHD Theorem (see also the end of Sect. 1.1.1 in Bonfiglioli and Fulci 2012).

With the above recollections at hand, we are ready to analyze more closely the contributions of Schur, Campbell, Poincaré, Pascal, Baker, Hausdorff, and Dynkin to the Exponential Theorem.

3 Schur's contribution

Schur's[8] papers (Schur 1890a,b, 1891, 1893) are to a large extent devoted to providing new derivations of Lie's Fundamental Theorems, using an insightful way of rewriting the group condition (3) via suitable systems of PDEs. Schur (1890a) provides conditions (involving the Jacobian matrix of the map $b \mapsto \varphi(a, b)$ at the identity) which turn out to be equivalent to some fundamental identities of Lie's theory. Actually, in dealing with a new proof of Lie's Third Fundamental Theorem, Schur (1890a, §3) constructs a transformation group (under a choice of canonical coordinates) only by means of suitable multivariate *power series* depending on the structure constants. *This is clearly a forerunner of the Campbell–Baker–Hausdorff Theorem.*

Indeed, anticipating the modern use of logarithmic coordinates (see Schur 1891, §1), Schur proves that the composition law $(a, b) \mapsto \varphi(a, b)$ related to the (canonically parametrized) group (7) can be written down explicitly as a series depending (polynomially) on the structure constants c_{ijk}. Schur, like Lie, did not catch the universal law of bracket-formation which underlies the analytic expression of $\varphi(a, b)$: hence his contribution to the form of the CBHD Formula as we now know it can be exaggerated. However, he perfectly realized that an explicit formula for φ, depending only on the c_{ijk}'s, was the right tool to prove Lie's Third Fundamental Theorem, a theorem which is, as we now understand, deeply related to the CBHD Theorem.

Schur's methods are direct and as natural as one should expect: besides the cited use of the canonical choice of parameters, they are mainly based on an integration of the above mentioned differentiated versions of the group property by the method of power series (the convergence of the resulting series being always carefully proved). In Schur (1890b), he then turns to represent the infinitesimal transformations of a transitive group by means of quotients of power series and of some absolute constants, structured on the *Bernoulli numbers* $\mathbf{B}_n$.

As an example of Schur's plain computations with power series, we would like to recall a remarkable formula (Schur 1890b, Eq. (2), p. 2), which catches one's eyes for its explicitness: If $\omega = (\omega_{i,j}(u))$ is the Jacobian matrix of $v \mapsto \varphi(u, v)$ at $v = 0$ (which gives the infinitesimal transformations for the so-called first parameter group), then

[8] Friedrich Heinrich Schur; Maciejewo, near Krotoschin, Prussia (now Krotoszyn, Poland), 1856—Breslau, Prussia (now Wrocław, Poland), 1932. Friedrich Schur should not be confused with the coeval mathematician, likely more famous, Issai Schur (Mogilev, 1875; Tel Aviv, 1941). While Friedrich Schur was a follower of the school of Sophus Lie, Issai Schur was a student of Frobenius and Fuchs in Berlin. For a source on Berlin mathematicians in the nineteenth century, see Biermann (1988).

$$\omega_{i,j}(u) = \delta_{i,j} + \sum_{m=1}^{\infty} \lambda_m \, U_{i,j}^{(m)}(u), \tag{12}$$

where $\delta_{i,j}$ is Kronecker's symbol, the constants λ_m are related to the Bernoulli numbers[9] by the relation $\lambda_m = \mathbf{B}_m/m!$, and the functions $U_{i,j}^{(m)}$ are explicitly defined by

$$U_{i,j}^{(m)}(u) = \sum_{\substack{1 \leqslant h_1,\dots,h_m \leqslant r \\ 1 \leqslant k_1,\dots,k_{m-1} \leqslant r}} c_{j,h_1}^{k_1} c_{k_1,h_2}^{k_2} \cdots c_{k_{m-2},h_{m-1}}^{k_{m-1}} c_{k_{m-1},h_m}^{i} \, u_{h_1} u_{h_2} \cdots u_{h_m}. \tag{13}$$

Here $u = (u_1, \dots, u_r)$ is a point of some neighborhood of the origin of $\mathbb{R}^r$ and the $c_{i,j}^s$ are the structure constants of the transformation group (in Schur's notation, $c_{i,j}^s$ stands for c_{ijs}; in the review to Schur (1890a), Engel firmly criticizes Schur for not generally following Lie's notations). Vice versa, given constants $c_{i,j}^s$ satisfying (10), i.e., the usual compatibility assumptions modeled on skew-symmetry and Jacobi identities, the above matrix $(\omega_{i,j})$ gives the component functions of the infinitesimal transformations of a transformation group having the $c_{i,j}^s$ as structure constants (see Schur 1891, Satz 2, p. 271).

Since a transformation group can be regained by exponentiating its infinitesimal transformations, the above results give information on the group itself. Aware of the impact of his results on the theory of transformation groups, Schur (1891) gives yet another simplified version of the results of his former papers and, later Schur (1893), the remarkable result that C^2 assumptions on the functions f and φ actually guarantee that they can be transformed (by a change of variables) into analytic functions. This very notable fact (besides opening the way to Hilbert's famous fifth problem[10]) has nowadays become a central topic of Lie group theory which is usually proved (by sheer chance) by means of the CBHD formula itself, highlighting the deep link of Schur's results with the Exponential Theorem.

While Schur will be mentioned by his successors mainly for the discovery of the role of the Bernoulli numbers in the composition of infinitesimal transformations, we explicitly remark that it will be Pascal who points out how the transformations in (12) can be obtained by the Exponential Theorem; see Pascal (1902c, §2, p. 565).

[9] Schur introduces the constants λ_m as satisfying a suitable recursion formula involving products of degree two (Schur 1890a, Eq. (56), p. 178):

$$(m+1)\lambda_m = -(\lambda_2\lambda_{m-2} + \lambda_3\lambda_{m-3} + \cdots + \lambda_{m-2}\lambda_2).$$

He then points out the relation of the λ_ms with the Bernoulli numbers (Schur 1890b, Eq. (4), p. 2) which he denotes as follows:

$$\frac{x}{e^x - 1} = 1 - \frac{x}{2} + \sum_{q=1}^{\infty} (-1)^{q-1} \frac{B_{2q-1}}{(2q)!} x^{2q}.$$

This means that, in our notation, $B_{2q-1} = (-1)^{q-1}\mathbf{B}_{2q}$.

[10] In fact, Hilbert quotes Schur (1893), see Alexandrov et al. (1979).

This is another confirmation of the intertwinement of Schur's studies with the CBHD Theorem.

4 Campbell's contribution

Campbell's[11] 1897 paper (Campbell 1897a) is the very first in the history of the Exponential Theorem. It can be considered as by far one of the most readable works on the subject: it is written in a concise style, yet its 10 pages contain a great variety of remarkable formulas which will return in every subsequent paper on the subject. Let us see what Campbell (1897a) proves.

To begin with, he establishes the following identity[12]:

$$\frac{y\,x^r}{r!}=\sum_{j=0}^{r}\frac{(-1)^j\,a_j}{(r+1-j)!}\sum_{i=0}^{r-j}x^i\,(\mathrm{Ad}\,x)^j(y)\,x^{r-j-i},\qquad r\in\mathbb{N}\cup\{0\},\tag{14}$$

where the constants a_j are defined by the recursion formula

$$\begin{cases} a_0=1, \quad a_1=1/2 \\ a_j=\frac{1}{j+1}\left(a_{j-1}-\sum_{i=1}^{j-1}a_i\,a_{j-i}\right), \quad j\geq 2. \end{cases}\tag{15}$$

Campbell acknowledges Schur for the discovery of the constants a_j (later in his second paper—Campbell (1898), he establishes the link between the a_j's and the Bernoulli numbers). Here x, y are pictured by Campbell as "operators," but actually they can be any elements of an associative non-commutative algebra.

Then, Campbell takes up an ingenious manipulation of certain formal power series, though neglecting any matter of convergence. Even if his intention is to apply his identities to the infinitesimal transformations of a group, his computations actually apply only in the algebra of the formal power series in two indeterminates x, y.

Though Campbell will eventually be faulted by Bourbaki for a lack of clarity about the context of his calculations (is it a transformation group, or the algebra of formal series, or a finite-dimensional algebra of vector fields?), the importance of his results on the Exponential Theorem is undoubted. For example, we feel compelled to exhibit Campbell's computation on p. 384 of Campbell (1897a) (which will reappear in modern proofs; see Cartier 1956 and Djoković 1975): Let us set

[11] John Edward Campbell; Lisburn (Ireland), 1862—Oxford (England), 1924.

[12] Here and in what follows, we use some different notations with respect to Campbell's, for those notations conflicting with the usual ones. For instance, Campbell denotes the sum $\sum_{i=0}^{r}x^i\,yx^{r-i}$ by $[yx^r]$ and he also introduces the sequence

$$y_1=yx-xy,\qquad y_r=y_{r-1}x-xy_{r-1}\quad(r\geq 2).$$

We preferred to reserve the square brackets for the commutator $[y,x]=yx-xy$, which Campbell instead denotes by (y,x), and to rewrite $y_r=[\cdots[[y,x],x]\cdots,x]$ in terms of left adjoints $(-\mathrm{Ad}\,x)^r(y)$.

$$z = z(x, y) := \sum_{j=0}^{\infty} (-1)^j a_j \, (\mathrm{Ad}\, x)^j (y) \tag{16}$$

(this will turn out to be the subseries of $\log(e^y e^x)$ containing y precisely once, a crucial series in the CBHD Formula). Then we have

$$\begin{aligned} y e^x &= \sum_{r=0}^{\infty} \frac{y\, x^r}{r!} \quad \text{(use (14) and interchange sums)} \\ &= \sum_{j=0}^{\infty} \sum_{r=j}^{\infty} \frac{(-1)^j a_j}{(r+1-j)!} \sum_{i=0}^{r-j} x^i \, (\mathrm{Ad}\, x)^j (y)\, x^{r-j-i} \quad \text{(rename } s = r - j) \\ &= \sum_{j=0}^{\infty} \sum_{s=0}^{\infty} \frac{(-1)^j a_j}{(s+1)!} \sum_{i=0}^{s} x^i \, (\mathrm{Ad}\, x)^j (y)\, x^{s-i} \quad \text{(sums can be interchanged)} \\ &= \sum_{s=0}^{\infty} \frac{1}{(s+1)!} \sum_{i=0}^{s} x^i \, z\, x^{s-i}. \end{aligned} \tag{17}$$

Since one has in general (as $\mu \to 0$)

$$(x + \mu z)^{s+1} = x^{s+1} + \mu \sum_{i=0}^{s} x^i \, z\, x^{s-i} + \mathcal{O}_s(\mu^2),$$

the above computation gives

$$\begin{aligned} (1 + \mu\, y)\, e^x &= e^x + \sum_{s=0}^{\infty} \frac{1}{(s+1)!} \Big((x + \mu z)^{s+1} - x^{s+1} + \mathcal{O}_s(\mu^2) \Big) \\ &= e^{x + \mu z} + \mathcal{O}(\mu^2). \end{aligned}$$

Note that this is the crucial point where the "symmetrized" sums $\sum_{i=0}^{s} x^i z x^{s-i}$ appear; analogous sums reappear in Poincaré (1900), Pascal (1901a), Baker (1905), and Hausdorff (1906).

Campbell's argument on the convergence of $\sum_{s=0}^{\infty} \mathcal{O}_s(\mu^2)$ is quite unconvincing (see Campbell 1897a, pp. 384–385), and this exhibits once again that his computations are to be considered only formally. However, Campbell obtains a fruitful formula:

$$\begin{cases} (1 + \mu y)\, e^x = e^{x + \mu z} + \mathcal{O}(\mu^2), \\ e^{\mu y}\, e^x = e^{x + \mu z} + \mathcal{O}(\mu^2), \end{cases} \quad \text{where } z = z(x, y) \text{ is as in (16)}, \tag{18}$$

and the a_j are explicitly given by (15).

A few years later, Poincaré will use similar formulas as a starting point for his ODE techniques in attacking the Exponential Theorem.

Now Campbell reveals to us his intent. He considers a set $\{X_1, \ldots, X_r\}$ of operators (i.e., linear analytic first-order differential operators on some domain of $\mathbb{R}^n$), such that $[X_i, X_j] = \sum_{k=1}^r c_{ijk} X_k$ (where the c_{ijk} are constants). Campbell aims to give a direct proof of Lie's Second Theorem and prove that, if X, Y belong to $V := \mathrm{span}\{X_1, \ldots, X_r\}$, then there exists $Z \in V$ such that $e^Z = e^Y \circ e^X$, where, to fix the notation, e^X is the finite transformation:

$$x \mapsto e^X(x) := \left(1 + X + \tfrac{1}{2!} X^2 + \cdots\right)(x).$$

In order to prove the Second Fundamental Theorem, his only tools are (18) and the following interesting fact. [Here we disambiguate Campbell's results by using some modern notation, obviously not used in Campbell (1897a).] Let us denote by $x \mapsto \gamma_X(t, x)$ the flow at time t of the vector field X starting at x (see also (6), (7), and (8)). Then one has (see Campbell 1897a, Eq. (β), p. 386)

$$\left(e^{tX} \circ Y \circ e^{-tX}\right)(x) = Y_{\gamma_X(t,x)}\,(\gamma_X(-t, \cdot)), \tag{19}$$

where $e^{\pm tX}$ on the left-hand side are considered as a series of differential operators. Now, the sides of (19) are as follows:

– after a simple computation, the left-hand side is

$$e^{t\,\mathrm{Ad}\,X}(Y) := \sum_{r=0}^{\infty} \frac{t^r}{r!}\,(\mathrm{Ad}\,X)^r(Y);$$

– the right-hand side is the operator Y'_x, obtained by expressing

$$Y_{x'} = \sum_{j=1}^{n} \eta_j(x')\,\frac{\partial}{\partial x'_j}$$

in the new coordinates x defined by $x = e^{-tX}(x') = \gamma_X(-t, x')$.

With Campbell's compact notation, (19) becomes[13]

$$e^{t\,\mathrm{Ad}\,X}(Y) = Y', \qquad \text{where} \quad \begin{cases} Y = \sum_{j=1}^{n} \eta_j(x)\,\dfrac{\partial}{\partial x_j} \quad \text{and} \\[2ex] Y' = \sum_{j=1}^{n} \eta_j(x'(x))\,\dfrac{\partial}{\partial x'_j}. \end{cases} \tag{20}$$

[13] Here $\partial/\partial x'_j$ must be properly interpreted as

$$\sum_{i=1}^{n} \frac{\partial \mathbf{x}_i}{\partial x'_j}(\mathbf{x}'(x))\,\frac{\partial}{\partial x_i},$$

where $\mathbf{x}(x') = \gamma_X(-t, x')$, and $\mathbf{x}'(x) = \gamma_X(t, x)$.

Anticipating some ideas from the subsequent paper (Campbell 1898), Campbell's proof of the Second Theorem could be described as follows[14]:

1. In determining $e^Y \circ e^X$, e^Y can be replaced by iterated applications of $1 + \mu\, Y$, where μ is so small that $\mathscr{O}(\mu^2)$ can be ignored. Indeed, as in ordinary algebra,
$$e^Y = \lim_{n\to\infty} (1 + Y/n)^n. \tag{21}$$
2. Thus, in computing $(e^Y \circ e^X)(x)$, it is enough to consider a recursion process, based on iterated applications of $1 + \mu\, Y$ applied to a finite transformation $x' = e^{\overline{X}}(x)$, where the infinitesimal transformation $\overline{X}$ will change at each step.
3. Now, by taking care of the proper substitutions, we have
$$\Big((1 + \mu Y) \circ e^X\Big)(x) = (1 + \mu\, Y')(x'),$$
where Y' has the same meaning as in the far right-hand side of (20). By the identity in the left-hand side of (20), we find that Y' is a Lie series in X, Y. Hence, Y' belongs to $V = \mathrm{span}\{X_1, \ldots, X_r\}$, since $X, Y \in V$, and V is a Lie algebra. Thus, the Second Fundamental Theorem will follow if we can write $(1 + \mu\, Y')(x')$ as $e^Z(x)$, where Y' is any element of V.
4. Given any X, Y in V, the first identity in (18) ensures that
$$(1 + \mu\, Y)\, e^X = e^{X + \mu\, Z_1} + \mathscr{O}(\mu^2),$$
where, by (16), $Z_1 = z(X, Y)$ is a series of brackets in X, Y. From the same reasoning as above, we infer $Z_1 \in V$. Setting $X_1 := X + \mu\, Z_1$, we have $X_1 \in V$ too, and the above identity can be rewritten as
$$(1 + \mu\, Y)\, e^X = e^{X_1} + \mathscr{O}(\mu^2).$$
5. By the same argument as in the previous step, we have
$$(1 + \mu\, Y)^2\, e^X = (1 + \mu\, Y)\, e^{X_1} + \mathscr{O}(\mu^2) \overset{(16)}{=} e^{X_2} + 2\, \mathscr{O}(\mu^2),$$
where $X_2 := X_1 + \mu\, Z_2$ and $Z_2 = z(X_1, Y)$.

As a matter of fact, Campbell considers the above arguments sufficient to obtain a proof of the Second Theorem and the paper (Campbell 1897a) simply ends with the suggestion to take $\mu = 1/n$, with n large. Probably aware of the fact that his previous argument is not convincing, Campbell writes a continuation paper (Campbell 1898). [Meanwhile, he also had to write a paper (Campbell 1897b) in correction of a statement made in Campbell (1897a) on a generalization which he sought to derive from his arguments.]

[14] As a matter of fact, Campbell's argument is not very transparent from now on, and his proof should be considered only as a sketch: nonetheless, we think it is interesting to see how he uses his previous remarkable computations.

Mainly, Campbell has to deal with the problems (evidently left open from the above arguments) to manage the "convergence" of the sequence X_n to some X_∞ and the vanishing of $n\,\mathscr{O}(\mu^2)$ as $n \to \infty$.

Let us briefly review the results of the second paper. First, Campbell recognizes the relationship between his constants a_j in (15) and the Bernoulli numbers.[15] Next, by means of more sophisticated computations on iterated bracket identities (valid in any associative algebra), he derives the following formal-series result[16]:

$$\begin{aligned}(\mathrm{Ad}\,(x+\mu z))^r\,(y) &= (\mathrm{Ad}\,x)^r \\ &+ \mu \cdot \sum_{m=0}^{r-1} \sum_{n=r-1-m}^{\infty} (-1)^{m+n-r}\, a_{m+n-r+1} \binom{r}{m} \Big[(\mathrm{Ad}\,x)^m(y), (\mathrm{Ad}\,x)^n(y)\Big] \\ &+ \mathscr{O}(\mu^2).\end{aligned} \tag{22}$$

This identity will help in dealing with a problem, contained implicitly in the previous paper, which we now describe in detail. In the following exposition, our task is to make precise Campbell's ideas in Campbell (1898, §13, p. 23), so where necessary we introduce some concepts and notation *not contained* therein (signaling our intervention in footnotes). To this aim, it is convenient to denote by $\mathfrak{R}$ the associative algebra of formal series in the letters x, y.[17] Following (16), we introduce a function

$$z : \mathfrak{R} \to \mathfrak{R}, \qquad z(u) := \sum_{j=0}^{\infty} (-1)^j a_j\, (\mathrm{Ad}\,u)^j\,(y),$$

[15] Campbell recognizes the relationship between his a_n and Schur's constants λ_n: it holds that $a_{2n+1} = 0$ for every $n \geq 1$ and

$$a_{2n} = (-1)^{n-1}\,\frac{B_{2n-1}}{(2n)!}, \quad n \geq 1,$$

if, as in Schur, the Bernoulli numbers B_{2n-1} are defined by

$$\frac{z}{e^z - 1} = 1 - \frac{z}{2} + \sum_{n=1}^{\infty} (-1)^{n-1}\, B_{2n-1}\, \frac{z^{2n}}{(2n)!}.$$

With our compact notation, we have $a_n = \frac{\mathbf{B}_n}{n!}$ for $n \geq 2$.

[16] See Campbell (1898, Eq. (D), p. 20), which must be corrected putting $(r - m)!$ in place of $(n - m)!$.

[17] As always, Campbell pictures x, y as differential operators, but he manipulates series in x and y with no care for convergence; hence, we preferred to introduce the algebra of formal power series in x and y.

where the a_j are as in (15).[18] Fixing a constant μ, let us consider the following sequence of maps $P_n : \mathfrak{R} \to \mathfrak{R}$, defined inductively by

$$\begin{cases} P_0(u) := u + \mu z(u), \quad P_1(u) := z(P_0(u)), \\ P_n(u) := P_{n-1}(P_0(u)), \quad n \geq 2. \end{cases} \tag{23}$$

In other words, denoting by P_0^n the n-fold composition of P_0 with itself, we have

$$P_n = z \circ P_0^n, \quad \text{for every } n \geq 1. \tag{24}$$

Next, we have to introduce another sequence of maps $x_n : \mathfrak{R} \to \mathfrak{R}$, namely

$$x_n(u) := P_0^n(u), \quad \text{for every } n \geq 0. \tag{25}$$

(Here it is understood that $P_0^0(u) = u$.) We shall soon be interested in $x_n(x)$: note that, because of (25) and the form of z, *every $x_n(x)$ is a Lie series in x, y universally determined by the Bernoulli-type numbers a_j.*

Since $P_0(u) = u + \mu z(u)$, we get

$$x_{n+1}(u) = x_n(u) + \mu z(x_n(u)), \quad \text{for every } n \geq 0. \tag{26}$$

As a simple inductive proof shows, the functions x_n can also be rewritten as follows:

$$x_n = P_0 + \mu\,(P_1 + P_2 + \cdots + P_{n-1}), \quad n \geq 2. \tag{27}$$

Now Campbell (1897a) recalls one of the crucial identities, namely, the first one in (18) above: treating the x as a pure symbol (allowing us to substitute for it any other element of our algebra $\mathfrak{R}$), one has

$$(1 + \mu y)\, e^u = e^{u+\mu z(u)} + \mu^2\, R(u), \quad \text{for every } u \in \mathfrak{R}. \tag{28}$$

Here $R(u)$ is some "remainder" term, to which Campbell does not pay too much attention. As a consequence of (26) and (28), we obtain at once

$$(1 + \mu y)\, e^{x_n(u)} = e^{x_{n+1}(u)} + \mu^2\, R(x_n(u)), \quad \text{for every } n \geq 0. \tag{29}$$

Applying $(1 + \mu y)^{r-n-1}$ to both members of (29), summing up for $n = 0, \dots, r-1$ and canceling out, Campbell provides the following important identity:

$$(1 + \mu y)^r\, e^x = e^{x_r(x)} + \mu^2 \sum_{n=0}^{r-1} (1 + \mu y)^{r-n-1}\, R(x_n(x)), \tag{30}$$

[18] Campbell never introduced z as a function, but, with long paraphrases, he repeatedly speaks of "substituting [some series] for x in z," when z "has been expressed explicitly in terms of x and y." The functional approach is evidently more convenient.

holding true for every $r \in \mathbb{N} \cup \{0\}$. The idea is now to take $\mu = 1/r$ and let $r \to \infty$. The above left-hand side then goes to $e^y e^x$ (see also (21)). Moreover, with a (maybe too) rough estimate on the second summand in the right-hand side of (30), Campbell states that this second summand can be ignored, for it has the same behavior as $r\ \mathcal{O}(1/r^2)$. He then concentrates on the existence (in a formal sense only) of

$$x_\infty := \lim_{r\to\infty} x_r(x),$$

and he attacks the problem of *deriving some explicit formula for* x_∞. To this aim, the decomposition (27) is employed, together with long and complicated estimates. For example, as far as the summands $P_0(x) + \mu\, P_1(x)$ are concerned, we have (note that the previous algebraic computations in (22) are now employed)

$$\begin{aligned}
P_0(x) + \mu\, P_1(x) &= x + \mu z(x) + \mu z(x + \mu z(x)) \\
&= \mu \sum_{j=0}^{\infty} (-1)^j a_j\, (\mathrm{Ad}\,(x + \mu z(x)))^j (y) \\
&= x + 2\mu z(x) + \mu^2 \cdot \left\{ \begin{array}{l} \text{certain universal sums of the constants } a_i a_j \\ \text{times brackets of the Lie elements } (\mathrm{Ad}\, x)^i (y) \end{array} \right\} + \mathcal{O}(\mu^3).
\end{aligned}$$

After several elaborate computations, Campbell derives a formula for x_∞ of the following form (note the ordering w.r.t. increasing powers of y rather than grouping homogeneous terms in x, y):

$$\begin{aligned}
e^y e^x &= e^{x_\infty} \quad \text{where} \\
x_\infty &= x + y + \frac{1}{2}[y, x] + \sum_{j=1}^{\infty} a_{2j}\, (\mathrm{Ad}\, x)^{2j} (y) \\
&\quad + \frac{1}{2!} \sum_{p,q=0}^{\infty} b_{p,q}\, \left[(\mathrm{Ad}\, x)^p (y), (\mathrm{Ad}\, x)^q (y)\right] \\
&\quad + \frac{1}{3!} \sum_{p,q,r=0}^{\infty} b_{p,q,r}\, \left[\left[(\mathrm{Ad}\, x)^p (y), (\mathrm{Ad}\, x)^q (y)\right], (\mathrm{Ad}\, x)^r (y)\right] + \cdots
\end{aligned} \tag{31}$$

where the constants $b_{p,q}, b_{p,q,r}, \ldots$ can be computed by a universal recursion formula, involving the numbers a_j in (15). For example,

$$b_{p,q} = (-1)^{p+q} \sum_{r=p+1}^{p+q+1} a_r\, a_{p+q+1-r} \binom{r}{p}.$$

Identity (31) can be considered by rights (even despite Campbell's not-completely-cogent derivation of it) as *the first form of the CBHD Theorem in the literature*.

We explicitly remark that, while in the first paper (Campbell 1897a) focussed on an Exponential Theorem within the setting of transformation groups, the second paper

(Campbell 1898) does not mention any underlying group structure: a remarkable break with respect to Campbell (1897a). Baker and Hausdorff will concentrate their attention in this abstract direction. Finally, Campbell (1903) provides a comprehensive treatise on transformation groups, much in the spirit of ODEs; in Chapter IV of his monograph (Campbell 1903), the proof of the Exponential Theorem (as given in Campbell 1897a) is reproduced in detail.

As we shall see in Sect. 6, Campbell's and Pascal's contributions are comparable from many points of view. Indeed, they share the following common features: many results (a symbolic Exponential Theorem with a sketched recursion formula for higher-order terms), some techniques, the laboriousness of the computations, the intents (applications to transformation group theory), the presence of some *lacunæ*. So the question is: why has Pascal's contribution been totally forgotten? The answer to this question, supposing it may ever be given, is lost in the twists and turns of mathematical history.

5 Poincaré's contribution

It was Poincaré[19] who first realized the link between Schur's proof of the Third Fundamental Theorem and what he defined "le Problème de Campbell" (this is the title of §IV in Poincaré (1900)).

Indeed, given a transformation group with infinitesimal transformations $X_1, \ldots, X_r$ as in (4), the one-parameter subgroups (see the exponential notation introduced in (8))

$$t \mapsto e^{tX} \qquad (X \in \mathfrak{g} := \operatorname{span}\{X_1, \ldots, X_r\})$$

generate the group of transformations. Hence, if Campbell's identity $e^V e^T = e^W$ has a precise analytic meaning (defining W as a *converging* Lie series in V and T, hence as an element of $\mathfrak{g}$), then it follows that the whole transformation group can be reconstructed from $\mathfrak{g}$ (or equivalently, by its structure constants). *Poincaré anticipated the modern way of thinking about the CBHD Theorem as the tool for reconstructing the group law by means of the brackets in the Lie algebra.* More precisely, as we shall explain later, he showed how to derive from Campbell's identity $e^V e^T = e^W$ a proof of the Third Fundamental Theorem, much in the spirit of Schur, as Poincaré himself declares: *C'est au fond la dèmonstration de Schur* (Poincaré 1900, p. 235).

The arguments sketched above show the necessity of giving a precise analytic solution to Campbell's problem, and Poincaré (1900) devotes the last five long paragraphs to this problem. We refer the reader to Schmid (1982) for a comprehensive account on Poincaré's contribution to the theory of Lie's transformation groups (and to Ton-That and Tran 1999 for his contribution to the Poincaré–Birkhoff–Witt Theorem); here we confine ourselves to pointing out his contribution to the development of the CBHD Theorem.

The most impressive fact about Poincaré's contribution is that, in solving an important problem, he creates an even more important tool: in fact, Sect. III of Poincaré

[19] Jules Henri Poicaré; Nancy (France), 1854—Paris (France), 1912.

(1900) gives birth to what we now call *the universal enveloping algebra* $\mathscr{U}(\mathfrak{g})$ of a Lie algebra $\mathfrak{g}$. Though Poincaré considers, ipso facto, only finite-dimensional Lie algebras over a field of characteristic zero, his ingenious idea is exactly the same one later rediscovered by Birkhoff and Witt in the general case. Suppose a finite-dimensional Lie algebra[20] is given, spanned by some "symbols" $X_1, \dots, X_r$ (called *generators*) and equipped with the bracket determined by the structure identities

$$[X_i, X_j] = \sum_{s=1}^{r} c_{ijs} X_s, \quad i, j = 1, \dots, r. \tag{32}$$

Poincaré then introduces an equivalence relation on the set of *formal polynomials* in the generators, by calling *equivalent* any two polynomials difference between which is a linear combination of two-sided products of the polynomials:

$$X_i X_j - X_j X_i - [X_i, X_j]. \tag{33}$$

After investigation of Poincaré (1900, §III), we agree with Hausdorff–Bourbaki in saying that Poincaré's exposition of the above construction is not free from some ambiguity.[21] Nonetheless, the cause for this ambiguity may be less trivial than it seems: Poincaré wanted to describe the way to construct $\mathscr{U}(\mathfrak{g})$ from $\mathfrak{g}$ for potentially arbitrary $\mathfrak{g}$'s; but then, in performing computations with infinitesimal transformations, he treats the formal product $XY - YX$ as the difference of true compositions of vector fields:

$$X \circ Y - Y \circ X.$$

Since (as we now know) the enveloping algebra of a finite-dimensional Lie algebra *of vector fields* $X_1, \dots, X_r$ can be identified with the associative algebra of the operators spanned by

$$X_1^{n_1} \circ \cdots \circ X_r^{n_r}, \qquad n_1, \dots, n_r \geq 0,$$

equipped with the composition of linear differential operators, Poincaré's identity

20 Lacking the general definition of a Lie algebra that was later introduced by Weyl in the 1930s, Poincaré pictured his Lie algebras as spanned by analytic vector fields: typically, the infinitesimal transformations of a group. However, in introducing the enveloping algebra, he never makes use of an underlying group of transformations.

21 For example, identity (1) on p. 224 writes $XY - YX = [X, Y]$ whereas identity (2) contains the non-trivial quantity $XY - YX - [X, Y]$, with the statement (a few lines below) that $XY - YX$ is of second degree whereas $[X, Y]$ is of first degree, whence they cannot be equal, contradicting the cited identity (1). As a matter of fact, Poincaré seems to write identity (1) only *to trace a parallel* between what happens for infinitesimal transformations of a group (where the bracketing has the alternant form $X \circ Y - Y \circ X$) and what it is expected to happen in the set of his formal polynomials, *modulo* (33). In other words, consistent with a later use of the symbol of equality in Poincaré (1900), the sign "=" in (1) should be interpreted as a congruence *modulo* the two-sided ideal generated by (33).

$$[X, Y] = XY - YX$$

can be applied with full justification in this vector field setting.

Next, Poincaré sketches the proof of the remarkable fact that any homogeneous formal polynomial of degree p is equivalent to a certain uniquely determined polynomial (called *regular*) which is a linear combination of polynomials of the form

$$(\alpha_1 X_1 + \alpha_2 X_2 + \cdots)^p.$$

In practice, this is the *(linear) isomorphism of* $\mathscr{U}(\mathfrak{g})$ *with the symmetric algebra of the vector space* $\mathfrak{g}$. We notice that he represents symmetric polynomials as powers of linear polynomials, a form particularly suited for the purpose of building up exponential series. This construction is exhaustively described (together with the filling of some *omissis* in Poincaré's arguments) in Ton-That and Tran (1999).

If V, T are linear combinations of $X_1, \ldots, X_r$, Poincaré considers a formal product of type:

$$e^V e^T = \sum_{m,n \geq 0} \frac{V^m T^n}{m!\, n!},$$

motivated by the theory of transformation groups.[22] His aim is to prove by direct methods what Lie's Second Fundamental Theorem ipso facto assures, namely, the existence of W such that $e^V e^T = e^W$. (Note that here Poincaré seems to return to the operator nature of these exponentials, rather than their being polynomial series.)

With his considerations on the universal enveloping algebra at hand, Poincaré then introduces a momentous way of rewriting $e^V e^T$, aiming directly at getting W, which deserves to be carefully considered for its originality. Indeed, the monomial $V^m T^n$ can be made equivalent to a sum of regular polynomials, say

$$\frac{V^m T^n}{m!\, n!} = \sum_{p=0}^{m+n} W(p, m, n), \tag{34}$$

[22] This is another delicate point: first Poincaré introduces e^{tX} as an *operator*, acting on functions in the obvious way

$$e^{tX}(f) = f + \frac{t}{1!} X(f) + \frac{t^2}{2!} X^2(f) + \frac{t^3}{3!} X^3(f) + \cdots,$$

having further care to recall that this series converges for small t (X is implicitly assumed to be analytic so that the above one is actually the expansion of $f(\gamma(t))$, where γ is an integral curve of X). But a few lines after, he writes $e^X e^Y = \sum_{m,n} \frac{X^m Y^n}{m!n!}$. If e^X, e^Y are considered as operators, then the correct expansion has X and Y interchanged. Consequently, Poincaré now considers e^X, e^Y only as *purely symbolic series*. This mixture of operator/symbolic objects (Poincaré often specializes his symbolic formulas in identities in the algebra of the operators) is the cause for some discrepancy.

where $W(p,m,n)$ (besides being symmetric) is homogeneous of degree p. Consequently, we get

$$e^V e^T = \sum_{p=0}^{\infty}\left(\sum_{m+n\geq p} W(p,m,n)\right) =: \sum_{p=0}^{\infty} W_p. \tag{35}$$

Note that W_p is a regular *formal series*, homogeneous of degree p. Poincaré observes that, if we seek for a W, homogenous of degree 1, such that $e^V e^T = \sum_{p=0}^{\infty} W^p$, then W^p is regular and homogeneous of degree p. Hence, by the *uniqueness* of the regularization process, we must have

$$W_p = \frac{(W_1)^p}{p!}, \quad p \geq 0. \tag{36}$$

Roughly speaking, W_1 *is nothing but the CBHD series.*

This fact deserves some clarification. In identity (34), we have

- V and T are elements of the Lie algebra $\mathfrak{g} := \mathrm{span}\{X_1,\ldots,X_r\}$ (recall that the X_j's satisfy (32));
- The "=" sign must be read as an identity in the enveloping algebra $\mathscr{U}(\mathfrak{g})$, i.e., a congruence modulo the ideal generated by (33). Thus, in computing the degree of homogeneity, the brackets $[X_i, X_j]$ must be considered of degree 1.

Alternatively, if we temporarily disregard the Lie algebra $\mathrm{span}\{X_1,\ldots,X_r\}$ and we consider only Lie algebras of vector fields (whose enveloping algebra can be identified with the associative algebra of polynomials in basis elements), then (34) may be considered as an authentic equality (not a congruence), obtainable by repeated applications of the identity:

$$VT = TV + [V,T]. \tag{37}$$

However, in this case, the concept of homogenous-degree is purely formal. This alternative use of V and T is not far-fetched, for two good reasons as follows:

- it reflects Poincaré's double use of the identity $[V,T] = VT - TV$ (later considered ambiguous by his reviewers);
- most importantly, *it allows the construction of the very important W_1 as a Lie series in the symbols V and T only, by means of a universal construction.*

For example, consider the following computation (where (37) is used):

$$\begin{aligned} VT &= \tfrac{1}{2}\,VT + \tfrac{1}{2}\,VT = \tfrac{1}{2}\,(VT+TV) + \tfrac{1}{2}\,[V,T] \\ &= \tfrac{1}{2}\,(V+T)^2 - \tfrac{1}{2}\,V^2 - \tfrac{1}{2}\,T^2 + \tfrac{1}{2}\,[V,T]. \end{aligned}$$

With the notation in (34), this gives

$$W(0,1,1) = 0, \quad W(1,1,1) = \tfrac{1}{2}\,[V,T], \quad W(2,1,1) = \tfrac{1}{2}\,\Big((V+T)^2 - V^2 - T^2\Big).$$

Again, consider the identity (which we can derive by means of (37) only)

$$\frac{VT^2}{2} = \tfrac{1}{6}(VTT + TVT + TTV) + \tfrac{1}{4}([V,T]T + T[V,T]) + \tfrac{1}{12}[T,[T,V]],$$

so that $W(1,1,2) = \frac{1}{12}[T,[T,V]]$. An analogous computation ensures that $W(1,2,1) = \frac{1}{12}[V,[V,T]]$, whence

$$\begin{aligned} W_1 &= W(1,1,0) + W(1,0,1) + W(1,1,1) + W(1,2,1) + W(1,1,2) + \cdots \\ &= V + T + \tfrac{1}{2}[V,T] + \tfrac{1}{12}[V,[V,T]] + \tfrac{1}{12}[T,[T,V]] + \cdots \end{aligned}$$

We recognize the first few terms of the CBHD series!

Since Poincaré is interested primarily in a new proof of the Third Fundamental Theorem (for which it is sufficient that W_1 be a convergent Lie series in the X_j's), an explicit derivation of a formula for W_1 in terms of repeated brackets of V and T is immaterial for the scope of Poincaré (1900), and is therefore omitted. But we agree with Schmid (1982, p. 182) who writes that "the answer follows easily from his methods." In the light of these facts, the original contribution by Poincaré to the CBHD is unquestionable.

As anticipated, Poincaré derives from (36) a new demonstration of the Third Fundamental Theorem of Lie. He clearly states that to form a transformation group with prescribed structure constants, the following chain of constructions suffices (Poincaré 1900, p. 234):

1. Given the structure constants, the process of reducing any polynomial into a symmetric one follows;
2. this allows for construction of the regular polynomials $W(p,m,n)$;
3. from the latter, one obtains W_1;
4. W_1 enables us to build up the functions $w_k = \varphi_k(v,t)$ such that $W = \sum_k w_k X_k$, $V = \sum_k v_k X_k$, $T = \sum_k t_k X_k$ do satisfy $e^V e^T = e^W$, a consequence of (36);
5. the map $v' = \varphi(v,t)$ provides the desired group.

It is interesting to have an overview of Poincaré's analytic solution of Campbell's Problem (36). A detailed description is beyond our scope here; we confine ourselves to exhibiting some formulas, which will eventually return in the literature devoted to the CBHD Formula.

Once again, Poincaré's ideas are destined to leave their marks: *he is the first to characterize the solution W of $e^W = e^V e^T$ as solving a suitable ODE, which he also writes down explicitly.* For instance, he proves that $e^V e^{\beta T} = e^{W(\beta)}$ if and only if $W(\beta)$ solves

$$\begin{cases} \frac{\mathrm{d}W(\beta)}{\mathrm{d}\beta} = \phi(\operatorname{Ad} W(\beta))(T) \\ W(0) = V \end{cases} \quad \text{where } \phi(z) = \frac{z}{1-e^{-z}}, \tag{38}$$

Poincaré (1900, equation (7) on p. 248). If V, T are elements of the Lie algebra $\operatorname{span}\{X_1, \ldots, X_r\}$ sufficiently near the origin, then $W := W(1)$ is the solution to

$e^W = e^V e^T$, an identity in the enveloping algebra of $\mathfrak{g}$. This identity translates into an identity between the operators e^V, e^T, e^W, which solves (at the same time) Campbell's Problem and the Second and Third Fundamental Theorems of Lie.

In many later proofs of the CBHD Theorem, formal power series of adjoint maps will frequently occur. But instead of analytically defining $\phi(\mathrm{Ad}\, W)$ by means of power series,[23] Poincaré prefers to invoke the residue calculus. This also gives him the possibility to make explicit that $W(\beta)$ is a linear combination of the X_i's. Indeed, by the cited use of the residue calculus, Poincaré is able to show that the (unique) solution $W(\beta)$ of (38) can be found in the form $\sum_{i=1}^r w_i(\beta)\, X_i$, where

$$\frac{\mathrm{d}w_i(\beta)}{\mathrm{d}\beta} = \frac{1}{2\pi\iota} \int \frac{\sum_{j=1}^r t_j\, P_{i,j}(\beta,\xi)}{\det(\mathrm{Ad}\,(W(\beta)) - \xi)} \, \frac{\xi}{1-e^{-\xi}}\, \mathrm{d}\xi, \tag{39}$$

and that these ODEs can be integrated. Here ι is the imaginary unit, the constants t_j are the coefficients of T with respect to $\{X_1, \ldots, X_r\}$, whereas $(P_{i,j})$ is the adjugate matrix (i.e., transpose of the cofactor) of $\mathrm{Ad}\,(W(\beta)) - \xi$, so that

$$(\mathrm{Ad}\,(W(\beta)) - \xi)^{-1} = \frac{\big(P_{i,j}(\beta,\xi)\big)_{i,j}}{\det(\mathrm{Ad}\,(W(\beta)) - \xi)}.$$

The integral is taken over a contour around $0 \in \mathbb{C}$ which does not contain the poles of $\phi(z)$, for example a circle about 0 of radius $R < 2\pi$. (Roughly speaking, the smallness of V, T reflects in the smallness of the eigenvalues of $\mathrm{Ad}\, W(\beta)$ up to $\beta = 1$, which in its turn ensures that the zeroes of $\det(\mathrm{Ad}\,(W(\beta)) - \xi)$ are surrounded by the cited contour.)

Poincaré arrives at his ODE formulation (38) of Campbell's Problem by using a mixture of different techniques: his process of symmetrization of polynomials in the enveloping algebra, differential calculus, explicit computations of non-commutative algebra, etc. For example, a key role is played by Eq. (7) in Poincaré (1900, p. 244), stating (in symbolic notation)

$$e^{\alpha V + \beta W} = e^{\alpha V} e^{\beta Y}, \quad \text{where } Y = \frac{1 - e^{-\mathrm{Ad}\,(\alpha V)}}{\mathrm{Ad}\,(\alpha V)}(W).$$

[23] Note that here we would have

$$\phi(\mathrm{Ad}\, W) = \sum_{n \geq 0} \frac{(-1)^n\, \mathbf{B}_n}{n!} (\mathrm{Ad}\, W)^n,$$

where the $\mathbf{B}_n$ are our Bernoulli numbers: $\mathbf{B}_n := \left(\frac{\mathrm{d}}{\mathrm{d}z}\right)^n \big|_0 \frac{z}{e^z - 1}$.

He derives this identity both with algebraic arguments and with a direct approach based on a β-expansion up to degree two:

$$
\begin{aligned}
e^{\alpha V+\beta W} &= e^{\alpha V} + \beta \sum_{n=1}^{\infty} \frac{\alpha^{n-1}}{n!}\left(\sum_{k=0}^{n-1} V^{n-1-k} W V^k\right) + \mathcal{O}(\beta^2)\\
&= e^{\alpha V} + \beta \sum_{n=1}^{\infty} \frac{\alpha^{n-1}}{n!}\left(\sum_{k=0}^{n-1} (-1)^k \tbinom{n}{k+1} V^{n-1-k} (\mathrm{Ad}\,V)^k(W)\right) + \mathcal{O}(\beta^2)\\
&= e^{\alpha V} + \beta \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \frac{1}{j!\,(k+1)!}(\alpha V)^j(-\alpha \mathrm{Ad}\,V)^k(W) + \mathcal{O}(\beta^2)\\
&= e^{\alpha V}\left(1 + \beta\,\frac{1-e^{-\mathrm{Ad}\,(\alpha V)}}{\mathrm{Ad}\,(\alpha V)}(W)\right) + \mathcal{O}(\beta^2).
\end{aligned} \tag{40}
$$

As we shall see, similar computations will return in Pascal, in Hausdorff and in plenty of modern proofs of the CBHD Theorem (see e.g., Djoković 1975).

It seems that Poincaré's contribution to the Exponential Theorem has shared the same fate as his contribution to the Birkhoff–Witt Theorem: both have been ignored for several decades. A full recovery of his idea in obtaining an explicit formula for the CBHD series from (34) would probably deserve a further investigation.

6 Pascal's contribution

In a series of five papers dated during 1901–1902 (and in the monograph Pascal 1903), Pascal[24] undertook the study of the composition of two exponentials, motivated by the search for new proofs of the Second and Third Fundamental Theorems of Lie. Here we give a brief overview of these papers, planning to return to this subject (with more mathematical content) in a forthcoming study.[25]

In Pascal (1901a), he announces a new proof of the Second Theorem of Lie: in his opinion, Lie's original proofs are too indirect to be considered definitive, and he expresses the aim to find the most direct proof as possible "*at the cost of longer computations*" (he will be eventually faulted by Hausdorff (1906), and by Engel reviewing Pascal 1902c, for these long calculations). Pascal understood from the beginning that the crucial tool was the study of the product of two finite transformations, under the canonical exponential form $x \mapsto e^{tX}(x)$.

With this aim in mind, Pascal (1901a) provides suitable identities occurring among the symbols of the finite transformations. Actually, all his computations are independent of any transformation group context and apply unaltered to indeterminates in any associative, non-commutative algebra. By means of the sole relation

[24] Ernesto Pascal; Naples (Italy), 1865–1940.

[25] Bonfiglioli, A.: The contribution of Ernesto Pascal to the so-called Campbell–Hausdorff formula, preprint.

$$[X_1, X_2] = X_1X_2 - X_2X_1,$$

he proves the following identity (see Eq. (15) in Pascal 1901a, p. 1067):

$$k\,X_2X_1^{k-1} = \sum_{j=0}^{k-1} j!\binom{k}{j}\gamma^{(j)}\left(\sum_{i=1}^{k-j} X_1^{i-1}\,(\mathrm{Ad}\,X_1)^j(X_2)\,X_1^{k-j-2}\right), \tag{41}$$

where $k \in \mathbb{N}$ and the numbers $\gamma^{(j)}$ are universal constants inductively defined by

$$\gamma^{(0)} = 1, \quad \gamma^{(j)} = -\left(\tfrac{1}{2!}\gamma^{(j-1)} + \tfrac{1}{3!}\gamma^{(j-2)} + \cdots + \tfrac{1}{j!}\gamma^{(1)} + \tfrac{1}{(j+1)!}\gamma^{(0)}\right). \tag{42}$$

These kinds of formulas involved in the Exponential Theorem were not new at the time. The same formulas appeared, a few years before, in Campbell (1897a, p. 381), but the scalar coefficients involved in this paper were defined by a very different recursion formula, so that Campbell's and Pascal's arguments are indeed independent. Moreover, we explicitly remark that the sum in parentheses in (41) belongs to the symmetric algebra of the vector space spanned by the brackets of X_1, X_2, so that (41) is actually a decomposition analogous to those studied by Poincaré (one year before Pascal) in defining the universal enveloping algebra.[26] Furthermore, Pascal's constants $\gamma^{(j)}$ are strongly related to the Bernoulli numbers, previously used by Schur, but Pascal's derivation of the role of the $\gamma^{(j)}$'s is clearly independent of Schur's.[27]

In his article (1901b), Pascal gives the application of his former symbolic identities to the study of the Second Fundamental Theorem of Lie. Indeed, he undertakes the study of an *explicit formula* for $e^{X_2} \circ e^{X_1}$, aiming to obtain the infinitesimal transformation X_3 generating this product. Pascal's point of view is very clear and incisive: in the opening of the paper, he says that a formula for X_3 provides, "*as a uniquely comprehensive source*," many applications to Lie group theory, such as the Second Theorem, the study of the permutability of one-parameter groups and the problem of the isomorphism of transformation groups (these applications are given in the subsequent paragraphs of Pascal 1901b).

[26] Pascal seemed to ignore, at the time, the paper by Poincaré (1900); it will be Engel, in his review of Pascal's subsequent paper (1901b), who points out the analogy with Poincaré's study. Admittedly, Pascal's technique is quite different from Poincare's, and his computations, though heavier, are more explicit.

[27] Schur's discovery of the role of Bernoulli numbers within the study of the Exponential Theorem will by cited by Pascal only after his paper (Pascal 1902a). It holds that, as Pascal will point out in the same article,

$$\gamma^{(0)} = 1, \quad \gamma^{(1)} = -\frac{1}{2}, \quad \gamma^{(2n+1)} = 0, \quad \gamma^{(2n)} = (-1)^{n+1}\frac{B_{2n}}{(2n)!},$$

where the numbers B_{2n} are defined by the following generating function:

$$\frac{x}{e^x - 1} = 1 - \frac{x}{2} + \sum_{n=1}^{\infty}(-1)^{n+1}\frac{B_{2n}}{(2n)!}x^{2n}.$$

With our notation for $\mathbf{B}_n$, Pascal's constants are nothing but $\gamma^{(n)} = \frac{\mathbf{B}_n}{n!}$—the same constants that appeared in Campbell but with a different recursion definition.

Some remarks are in order. Pascal's awareness of the reach of his result is enthusiastically affirmed, but he fails to refer to Campbell's earlier investigations and, mostly, to those by Poincaré. The analogy with Poincaré's point of view (recognizing the Exponential Theorem as a unifying tool) is evident, but there is no way of knowing if Pascal knew Poincaré's paper (1900) at the time.

It will be Engel in his review of Pascal (1901b), who points out the analogy with Poincaré (1900) and, most importantly, underlines a fault in Pascal's paper: the convergence of the series expressing X_3 in terms of X_1, X_2 is lacking. This makes it impossible to apply Pascal's result directly. Even if Engel's words sound quite hard,[28] he acknowledges that Pascal's formula is very remarkable.

Therefore, let us consider Pascal's argument in detail. He decomposes the product

$$\frac{X_2^r}{r!}\,\frac{X_1^{k-r}}{(k-r)!} \qquad (0 \le r \le k,\ k \in \mathbb{N})$$

as a linear combination of certain symmetric sums (called *elementary* of order $k, k-1, k-2, \ldots$) based on X_1, X_2 and scalar products of nested commutators of the form:

$$[X_{i_1} X_{i_2} \cdots X_{i_s} X_1 X_2] := [X_{i_1}, [X_{i_2} \cdots [X_{i_s}, [X_1, X_2]] \cdots]], \tag{43}$$

where $i_1, \ldots, i_s \in \{1, 2\}$. The law of composition of the scalar coefficient of such an elementary sum is described very closely, though an explicit formula is not given. Pascal states that the cited decomposition of $\frac{X_2^r}{r!}\frac{X_1^{k-r}}{(k-r)!}$ can be obtained by an inductive argument deriving from the identities in his former paper (1901a). (Actually, we think that this point deserves more detail.)

With this decomposition at hand, he rewrites the product:

$$\sum_{i,j\ge 0} \frac{(t'X_2)^j}{j!}\,\frac{(tX_1)^i}{i!} = \sum_{k\ge 0}\left(\sum_{r=0}^{k} t'^r\, t^{k-r}\, \frac{X_2^r}{r!}\,\frac{X_1^{k-r}}{(k-r)!}\right).$$

For instance, his "semi-explicit" form of the coefficients of the elementary sums allows him to illustrate that the above sums in parentheses reconstruct into a pure exponential:

$$\sum_{k=0}^{\infty} \frac{1}{k!}\left(tX_1 + t'X_2 + \gamma^{(1)}\, tt'\,[X_1, X_2] + \cdots\right)^k,$$

[28] Engel says that Pascal's application to the Second Fundamental Theorem is "devoid of any foundation" if the proof of the convergence is missing, and he faults Pascal for "surprisingly omitting to ask himself the question." Engel also reveals that Lie himself had calculated the first six terms of this series, but then he gave up the study of this series for "he did not foresee any means of answering the question of convergence."

where the coefficients $\gamma^{(j)}$ are as in (42). Note the similarity with Poincaré's decomposition (resulting from (35) and (36)):

$$e^V e^T = \sum_{p=0}^{\infty} \frac{(W_1)^p}{p!}, \quad \text{with } W_1 = W(1,1,0) + W(1,0,1) + W(1,1,1) + \cdots. \tag{44}$$

Exactly like Poincaré, when he stated that W_1 can be constructed by means of the reduction of $\frac{V^m T^n}{m!n!}$ to a regular polynomial, Pascal writes that the determination of X_3 follows by means of his universal decomposition (treated in Pascal 1901a and Pascal 1901b, §1). Nonetheless, whereas Poincaré uses an a posteriori argument to prove equality (44) (by showing that W_1 satisfies a suitable system of ODEs), Pascal's construction is much more direct: as he announced, at the cost of onerous computations, he provides a direct way to reconstruct the product $e^{t'X_2}e^{tX_1}$ as an exponential e^{X_3}, this reconstruction being uniquely based on an unraveling of $\sum_{i,j} \frac{(t'X_2)^j}{j!} \frac{(tX_1)^i}{i!}$.

For instance, his method permits him to exhibit the following terms:

$$\begin{aligned} X_3 = {} & tX_1 + t'X_2 + \gamma^{(1)}tt'\,[X_1, X_2] \\ & +\gamma^{(2)}t^2t'\,[X_1X_1X_2] - \gamma^{(2)}tt'^2\,[X_2X_1X_2] - \gamma^{(2)}\gamma^{(1)}t^2t'^2\,[X_1X_2X_1X_2] \\ & +\gamma^{(3)}t^3t'[X_1X_1X_1X_2] - \gamma^{(3)}tt'^3[X_2X_2X_1X_2] + \cdots \end{aligned} \tag{45}$$

Since $\gamma^{(1)} = -1/2$, $\gamma^{(2)} = 1/12$, and $\gamma^{(3)} = 0$, we recognize the first few terms of the CBHD series. Besides this partial expansion, Pascal proves that X_3 is a series of terms as in (43). As we remarked, *Poincaré did not manage to prove this result explicitly.* It was Campbell who stated a similar fact, but with a completely different technique.

Finally, in Pascal (1901b, §2) Pascal gives the following proof of Lie's Second Theorem. (As Engel pointed out, the convergence of the series for X_3 is needed to make the argument valid.) Let $Z_1, \dots, Z_r$ be infinitesimal transformations, and set $V := \mathrm{span}\{Z_1, \dots, Z_r\}$; he aims to find the necessary and sufficient conditions for the above X_3 to belong to V, whenever X_1, X_2 do. Obviously, if $[Z_h, Z_k] \in V$ for every h, k, then $X_1, X_2 \in V$ implies $X_3 \in V$, since the summands in X_3 are all of the form (43). Vice versa, if $X_1, X_2, X_3 \in V$, then $X_3 - tX_1 - tX_2 \in V$ for every (small) t. By the expansion in (45) up to degree 2 in X_1, X_2 jointly, this gives

$$V \ni \gamma^{(1)}t^2\,[X_1, X_2] + \mathcal{O}_{t\to 0}(t^3).$$

If we divide by t^2 and we let $t \to 0$, then we get $[X_1, X_2] \in V$. Since we can take any Z_h, Z_k in place of X_1, X_2, this gives $[Z_h, Z_k] \in V$, and Lie's Second Fundamental Theorem follows.

The paper (Pascal 1902a) is completely devoted to some first and second degree identities concerning the Bernoulli numbers. With a considerable amount of computations (showing his expertise in handling special functions), Pascal provides identities involving the numbers $\gamma^{(j)}$ and the products $\gamma^{(i)}\gamma^{(j)}$. These identities will be helpful in his new concern: to prove Lie's Third Theorem with a direct approach, similar to the one that he followed to prove the Second Theorem. At last, in the *incipit* of the paper, Pascal cites Schur (1890a) (for the discovery of the roles of Bernoulli numbers in the study of the infinitesimal transformations of the so-called parameter group) and Campbell (1897a, 1898) for "certain formulas on the infinitesimal transformations, derived with different methods." He seems not to attribute to Campbell any effective discovery on the Exponential Theorem (this was the same attitude as Poincaré's and Hausdorff's).

A certain reluctance in quoting other authors is manifestly shown in the way Pascal mentions the contributions by Schur and by Poincaré to the Third Theorem: for example, he simply says that Poincaré had "recently studied this problem," but no reference of any work is provided. [It is evident that, even supposing that Pascal got to know Poincaré's work only after the article (Pascal (1901b)), a comparison with (Poincaré (1900)) was now obligatory.]

The study of Lie's Third Theorem is taken up in Pascal (1902b). This time the quotation of previous works is more accurate: he mentions Schur's contributions (1890a; 1890b; 1893) as "extended and important;" he also briefly quotes Poincaré's recent studies (1899; 1900) as "starting from different points of view."

The paper is devoted to explicitly constructing the infinitesimal transformations *of a transformation group*, provided constants c_{ijk} are given, satisfying (10). He acknowledges the similarity with Schur's papers, but he stresses that his own method "is based to a large extent on some identities on Bernoulli numbers" as provided in the article (Pascal 1902a), "without any notion of group theory other than the Second Fundamental Theorem." (We agree with Pascal in observing that, even if Schur's methods could be considered direct as well, they require a more overt use of the theory of transformation groups and ODEs.)

After two long sections of laborious computations of identities involving the constants c_{ijk} and $\gamma^{(j)}$, the proof of the Third Theorem is given as follows. Let $u_1, \ldots, u_r$ be coordinates on some neighborhood of 0. Consider the following vector field:

$$U_k := \sum_{h=1}^{r} \Bigg(\delta_{hk} - \gamma^{(1)} \sum_{t_1} c_{t_1 kh} u_{t_1} + \sum_{n=1}^{\infty} \gamma^{(2n)} \sum_{t_1, t_2, \ldots} \times \sum_{s_1, s_2, \ldots} c_{t_1 k s_1} c_{t_2 s_1 s_2} c_{t_3 s_2 s_3} \cdots c_{t_{2n} s_{2n-1} h} \, u_{t_1} \cdots u_{t_{2n}} \Bigg) \frac{\partial}{\partial u_h}, \tag{46}$$

where δ_{hk} is Kronecker's symbol (the above is Eq. (16) in Pascal 1902b, p. 428). Then Pascal proves (this time with a very accurate estimate) that the above power series converges on some neighborhood of 0, so that $U_1, \ldots, U_r$ are well defined. Furthermore, the necessary identity

$$[U_i, U_j] = \sum_{k=1}^{r} c_{ijk} \, U_k \tag{47}$$

holds true, so that (by Lie's Second Theorem) the infinitesimal transformations U_k generate a group of transformations with structure constants c_{ijk}. Identity (47) is proved by means of the computations of the previous sections of the paper, together with those in another (Pascal 1902a). We explicitly note the similarity with the results obtained by Schur: see (12) and (13).

Pascal returns to the above argument in Pascal (1902c). Probably aware of the similarity of his results if compared to Schur's, in Pascal (1902c) he seizes the opportunity to show that the vector fields in (46) can be derived from his own previous purely algebraic results. Accidentally, he also exhibits the fact that Schur's explicit formulas can be derived by the Exponential Theorem, thus unveiling the precursory merits of Schur's study for the later CBHD Theorem.

Indeed, the paper (Pascal 1902c) must be considered as the *summa* of Pascal's works on the theory of transformation group (later collected in the monograph Pascal 1903), because:

- he only uses his original algebraic identities in Pascal (1901a) (in neat difference with Schur's methods);
- by means of these identities, he provides the explicit expression of suitable coefficients in the expansion (45), thus improving the results in Pascal (1901b) (as Hausdorff will later discover with different techniques, these crucial coefficients are those appearing in Campbell's series (16), and Pascal gives the expression of precisely these ones);
- he constructs, with a more natural method when compared to Pascal (1902b), the infinitesimal transformations of the so-called parameter group (this time using some group prerequisites).

The paper (Pascal 1902c) contains the most important of Pascal's contributions to the CBHD Theorem, and (as we shall discuss below) many of his computations will return as crucial tools in Baker and in Hausdorff (but without any acknowledgement of this fact by these authors).

Therefore, let us briefly give an overview of the key results in Pascal (1902c). In Sect. 1 Pascal corrects a mistake from his paper (1902b) (the same mistake occurred also in Poincaré (1900); it will return again in Baker (1905)): If T, S are, respectively, the finite transformations

$$\begin{cases} T: & x' = e^{t'X_2}(x) \\ S: & x'' = e^{tX_1}(x'), \end{cases}$$

then the composition $S \circ T$ of S *after* T admits the correct expansion:

$$e^{tX_1} \circ e^{t'X_2} = \sum_{r,s \geq 0} t'^s t^r \frac{X_2^s X_1^r}{s!\,r!}.$$

This shows that some care must be taken in passing from finite transformations of exponential type to the symbolic exponential calculus.

He then shows that, in the expansion (45), all the summands containing t' with degree 1 (and t with any degree ≥ 1) are given by

$$\sum_{n=1}^{\infty} \gamma^{(n)} t' t^n \, (\mathrm{Ad}\, X_1)^n (X_2), \tag{48}$$

where the constants $\gamma^{(n)}$ are as in (42). Analogously, the summands containing t with degree 1 (and t' with any degree ≥ 2) are given by

$$-\sum_{n=2}^{\infty} \gamma^{(n)} t'^n t \, (\mathrm{Ad}\, X_2)^{n-1}([X_1, X_2]) = \sum_{n=2}^{\infty} \gamma^{(n)} t'^n t \, (\mathrm{Ad}\, X_2)^n (X_1). \tag{49}$$

Let now $Z_1, \dots, Z_r$ be the infinitesimal transformations generating a transformation group with r parameters. Let us set

$$X_1 = v_1 Z_1 + \cdots + v_r Z_r, \qquad X_2 = u_1 Z_1 + \cdots + u_r Z_r.$$

Then, by general transformation group theory, it is known that the composition $e^{X_1} \circ e^{X_2}$ is given by e^{X_3}, where X_3 is a linear combination of the Z_i's, say

$$X_3 = u'_1 Z_1 + \cdots + u'_r Z_r.$$

The coefficients u'_i are functions of u and v (and of the structure constants), say

$$u'_h = \varphi_h(u, v), \qquad h = 1, \dots, r. \tag{50}$$

Again from transformation group theory, it is known that (50) defines a finite group of transformations, called the *parameter group*.

Now, thanks to the formula for the product of two exponentials, one can find all of the functions φ_h. Pascal remarks that, if we know that (50) defines a transformation group, it is sufficient to consider only the terms containing v with degree one (which furnish, ipso facto, the infinitesimal transformations): these correspond to the summands in (45) with t having degree one. By formula (49) (plus a summand from (48)) we get[29]

$$\begin{aligned} X_3 = {} & \sum_i v_i Z_i + \sum_i u_i Z_i + \gamma^{(1)} \left[\sum_i v_i Z_i, \sum_j u_j Z_j \right] \\ & + \sum_{n=2}^{\infty} \gamma^{(n)} \left(\mathrm{Ad} \sum_i u_i Z_i \right)^n \left(\sum_j v_j Z_j \right) + \mathcal{O}(|v|^2). \end{aligned}$$

[29] Once again, a discussion on the convergence of this series would be required, but Pascal omits it.

By Lie's Second Fundamental Theorem, it holds that $[Z_i, Z_j] = \sum_k c_{ijk} Z_k$, whence

$$(\mathrm{Ad}\, X_2)^n(X_1) = \sum_{t_1,\dots,t_n} \sum_{s_1,\dots,s_{n-1}} \sum_{h,k} c_{t_1ks_1} c_{t_2s_1s_2} \cdots c_{t_ns_{n-1}h}\, u_{t_1} \cdots u_{t_n} v_k\, Z_h.$$

As a consequence we derive

$$\varphi_h(u,v) = u_h + v_h + \frac{1}{2} \sum_{jk} c_{jkh}\, u_j v_k + \sum_{n=1}^{\infty} \sum_{t_1,\dots,t_{2n}} \sum_{s_1,\dots,s_{2n-1}} \sum_k \times \gamma^{(2n)}\, c_{t_1ks_1} \cdots c_{t_{2n}s_{2n-1}h}\, u_{t_1} \cdots u_{t_{2n}} v_k + \mathscr{O}(|v|^2).$$

Differentiating this with respect to the v's, Pascal derives at once

$$U_k = \sum_{h=1}^{r} \frac{\partial \varphi_h}{\partial v_k}(u,0) \frac{\partial}{\partial u_h} = \sum_{h=1}^{r} \Bigg(\delta_{hk} + \frac{1}{2} \sum_j c_{jkh} u_j + + \sum_{n=1}^{\infty} \gamma^{(2n)} \sum_{t_1,t_2,\dots} \sum_{s_1,s_2,\dots} c_{t_1ks_1} \cdots c_{t_{2n}s_{2n-1}h}\, u_{t_1} \cdots u_{t_{2n}} \Bigg) \frac{\partial}{\partial u_h},$$

which turns out to be exactly formula (46) of his previous work (Pascal 1902b) (and the analogue of Schur's (12)). Finally, another group theoretic result guarantees that the full expression of the finite transformation φ can be regained by exponentiation:

$$\varphi(u,v) = e^{\sum_k v_k U_k}(u) = u + \Big(\sum_k v_k U_k\Big)(u) + \frac{1}{2!}\Big(\sum_k v_k U_k\Big)^2 (u) + \cdots$$

In view of the above explicit formula for the infinitesimal transformations U_k, this identity contains a very "quantitative" version of the Third Fundamental Theorem, in that it shows that an explicit transformation group can be constructed by the use of the Bernoulli numbers $\gamma^{(2n)}$ and by a set of constants c_{ijk} satisfying the structure relations (10). This fact was not new at the time, in view of Schur's former studies, but the way it is derived is indeed new: this is done by the use of Pascal's version of the Exponential Theorem.

Pascal's contribution to the development of the CBHD Theorem is paramount: in modern parlance, we can say that he has shown for the first time how to construct an explicit (transformation) group by using the CBHD series

$$X \diamond Y = X + Y + \tfrac{1}{2}[X,Y] + \tfrac{1}{12}[X,[X,Y]] - \tfrac{1}{12}[Y,[X,Y]] + \cdots$$

associated to X and Y, or, more precisely, by using the subseries extracted from $X \diamond Y$ of the summands of degree one in one of the indeterminates X or Y. Analogous results will be re-obtained by Baker and, mostly, by Hausdorff, and will reappear in more modern proofs of the CBHD Theorem; see Reutenauer (1993, Sect. 3.4).

7 Baker's contribution

Preceded by a series of papers (Baker 1901, 1902, 1903) mostly concerned with exponential-type theorems in the context of transformation groups and groups of matrices, Baker[30] devoted an important paper (1905) to the Exponential Theorem for abstract non-commutative indeterminates.

His words speak very clearly: "It is an obvious suggestion of Lie's theory of groups that the product $e^A e^B$ is of the form e^C, where C is a series of alternants (i.e., brackets) of A and B." His aim is to prove this fact by a method "of the kind usually called symbolical," in order to provide "a contribution to the calculus of alternants of any non-commutative quantities."

As a connection, he cites Campbell's proof in Campbell (1898, 1903) as "an elementary proof" of the symbolic theorem, though he considers this proof as not final. Later, he provides an application of his Exponential Theorem to Lie's Second Theorem, Baker (1905, p. 44) saying that an analogous approach was "in view" in Campbell's paper (1898). Nothing is said about Campbell's technique and no mention at all is made of Poincaré's related results nor of Pascal's. Schur is luckier; he is acknowledged for his study of the parameter group and (see Baker 1905, p. 26) for his explicit formulas exhibiting a transformation group with given structure constants. Baker states that by means of his own Exponential Theorem, he can recover Schur's explicit group. Actually, this fact was already pointed out and proved in detail by Pascal (but Baker gives no reference of this fact, probably being unaware of that paper Pascal 1902c).

Baker's words celebrate the Exponential Theorem (incidentally, he is seemingly the first one to use this expression; see the title of his paper (1901) and that of §3 in his paper (1905)) as a universal tool in giving "the parameter group of every possible continuous group." As a matter of fact, he was preceded in this intuition by all his predecessors: Schur (implicitly), Poincaré, Pascal, and Campbell.

The first palpable novelty of Baker's approach is to allow consideration of symbolical expressions of the form:

$$(e^A e^B - 1) - \frac{1}{2}(e^A e^B - 1)^2 + \frac{1}{3}(e^A e^B - 1)^3 - \cdots \tag{51}$$

This is the first time that a genuine *logarithm*, namely $\log(e^A e^B - 1)$, makes its explicit appearance in the proof of the Exponential Theorem. This is possible because Baker considers A and B as pure "non-commutative quantities," as he defines them. In the second page of his paper, we also find the explicit expansion of the above series up to the fourth order[31]:

[30] Henry Frederick Baker; Cambridge (England), 1866–1956.

[31] We henceforth denote Baker's alternants (A, B) with the usual $[A, B]$; also, although Baker never used this notation, we also allow ourselves to use adjoint maps like Ad A, Ad B, ... with the usual meaning.

$$A + B + \frac{1}{2}[A, B] + \frac{1}{12}[A - B, [A, B]] - \frac{1}{24}[A, [B, [A, B]]]. \tag{52}$$

Then, Sect. 1 is devoted to introducing a certain formalism concerned with bracketing. At a first reading this formalism looks puzzling and mystifying.[32] Baker intends it to simplify computations with brackets, in that it comprises many identities resulting from skew-symmetry and Jacobi identities. Let us take a look at this formalism, which Baker describes rather than rigorously defines.

If capitals $A, B, C, \dots$ denote elements of some associative but non-commutative algebra,[33] Baker associates to each of them a lower-case $a, b, c, \dots$ The pairing $A \leftrightarrow a$ is injective and linearly extended, and a is called the *base* of A, whereas A is called the *derivative* of a. He uses the notation ϵ to mean the map defined by

$$\epsilon a := A, \quad \text{where } a \text{ is the base of } A.$$

He then proposes to define the base of any element in the *Lie algebra generated by capitals*. He begins with nested brackets (called "simple alternants"): the base of $[A, B]$ is denoted by Ab, whence skew-symmetry imposes the identities

$$Ab + Ba = 0, \quad Aa = 0. \tag{53}$$

Moreover, if $d = Bc$ is the base of $[B, C]$, the base of $[A, [B, C]]$ has to be Ad, also denoted by ABc. Accordingly,

$$A_1 A_2 \dots A_n b \quad \text{denotes the base of} \quad [A_1, [A_2 \cdots [A_n, B] \cdots]].$$

This notation is not ambiguous because of the associativity of composing adjoint maps, which Baker summarizes by saying that in the product $A_1 A_2 \dots A_n b$ "the symbols are associative." The next task is to define the base of a non-simple bracket (called "compound alternant"), for example of $D := [[A, B], C]$. Since, by skew-symmetry, this equals $-[C, [A, B]]$, we expect d to be $-CAb$. But the Jacobi identity also imposes that $D = [A, [B, C]] + [B, [C, A]]$, so that we must also have $d = ABc + BCa$. This provides the identity, which substitutes for Jacobi's,

$$ABc + BCa + CAb = 0. \tag{54}$$

[32] In reviewing Baker (1905), Engel says that this formalism has the effect of making "a little heavier the understanding" of the contents; then, with harder words, Engel says that Baker's use of bases and derivatives is definitely a *ziemlich überflüssige Einführung, a quite superfluous introduction.*

[33] Apparently, Baker fixes some letters defining what we would call an algebra of words, which is associative but not commutative. He never clarifies which is the basis of this algebra, so that his further constructions may be in contrast with possible linear dependence relations.

Baker goes on with other delicate identities concerning bases and derivatives, proving that his symbols obey some natural associative and distributive laws.[34] As he declares, the use of bases and derivatives "furnishes a compendious way of expressing the relations among alternants." In fact, as (53)–(54) show, Baker's symbolic technicalities summarize the basic relations existing in a Lie algebra and furnish new ones: for example, from (53) with $A = [B, C]$, one gets $[B, C]Bc = 0$; the associative and distributive laws then give $BCBc - CB^2c = 0$; finally, by applying the map ϵ, we derive

$$[B, [C, [B, C]]] - [C, [B, [B, C]]] = 0.$$

Furthermore, Baker also allows the basic laws of algebra to continue to hold for formal power series constructed with his symbols. For example, the derivative of $c = \left(\sum_{i=0}^{\infty} \lambda_i \, A^i\right) b$ is

$$C = \sum_{i=0}^{\infty} \lambda_i \, (\mathrm{Ad}\, A)^i(B),$$

and if $\lambda_0 \neq 0$ and $\sum_{i=0}^{\infty} \mu_i \, z^i$ is the formal power series reciprocal to the series $\sum_{i=0}^{\infty} \lambda_i \, z^i$, then $b = \left(\sum_{i=0}^{\infty} \mu_i \, A^i\right) c$ so that

$$C = \sum_{i=0}^{\infty} \mu_i \, (\mathrm{Ad}\, A)^i(B).$$

A more delicate topic is the so-called "substitutional operation" treated in Sect. 2. This same operation will return in Hausdorff's paper with no mention of Baker's prior study (of which Hausdorff was probably unaware). Let us summarize it.

Given bases A, B with derivatives a, b, the symbol $b \frac{\partial}{\partial a}$ defines the operation replacing a and A by b and B (respectively) *one at a time*. For example

$$\left(b \frac{\partial}{\partial a}\right) A^2Ca = BACa + ABCa + A^2Cb.$$

In practice, Baker is defining a sort of mixed *algebra-derivation*, operating both on bases and derivatives.[35] As a first main result on substitutions, he takes any $B = \left(\sum_{i=0}^{\infty} \lambda_i \, A^i\right) C$, and he considers the formal power series in the real number t given by

[34] For example, he provides a "distributive law": Let $A = [A_1, \cdots [A_{n-1}, A_n] \cdots]$ be expanded in the form, $A = \sum \pm K_1 \ldots K_n$ (where the K's are permutations of the A's). Then he proves that

$$[A, B] = \left[\sum \pm K_1 \ldots K_n, B\right] = \sum \pm [K_1, [K_2 \cdots [K_n, B] \cdots]],$$

so that $\left(\sum \pm K_1 \ldots K_n\right) b = \sum \pm (K_1 \ldots K_n b)$.

[35] Since it is never made clear which capitals are generators of the whole algebra of capital letters, the well-posedness of this derivation is not so obvious.

$$F(t) := \sum_{i=0}^{\infty} \frac{t^i}{i!} \left(b\, \frac{\partial}{\partial a} \right)^i A.$$

Then, it is easily seen that one has

$$\frac{\mathrm{d}F}{\mathrm{d}t}(t) = \left(b\, \frac{\partial}{\partial a} \right) F(t). \tag{55}$$

Now let us also put, when F is as above,

$$G(t) := \sum_{j=0}^{\infty} \mu_j\, F(t)^j.$$

Then, by means of (55), it is not difficult to see that (setting $\delta := b\, \frac{\partial}{\partial a}$ for brevity)

$$\frac{\mathrm{d}^m G}{\mathrm{d}t^m}(0) = (\delta^m G)(0) = \delta^m(G(0)).$$

As a consequence, by expanding G in powers of t we get (see Baker 1905, p. 33)

$$G(t) = \sum_{i=0}^{\infty} \frac{t^i}{i!} \left(b\, \frac{\partial}{\partial a} \right)^i G(0), \qquad \text{with } G(0) = \sum_{j=0}^{\infty} \mu_j\, A^j, \tag{56}$$

i.e., more compactly,

$$\sum_{i=0}^{\infty} \frac{t^i}{i!} \left(b\, \frac{\partial}{\partial a} \right)^i \left(\sum_{j=0}^{\infty} \mu_j\, A^j \right) = \sum_{j=0}^{\infty} \mu_j \left\{ \sum_{i=0}^{\infty} \frac{t^i}{i!} \left(b\, \frac{\partial}{\partial a} \right)^i A \right\}^j, \tag{57}$$

where b is any base of the form $b = \left(\sum_{i=0}^{\infty} \lambda_i\, A^i\right) c$. Identity (57) can be interpreted, in modern parlance, as follows: *Since* $\delta = b\, \frac{\partial}{\partial a}$ *is a derivation,* $\exp(t\delta) := \sum_{i \geq 0} t^i \delta^i / i!$ *is an algebra morphism, and* (57) *follows by "continuity."*

Another very interesting identity is the following one (see Baker 1905, p. 34):

$$\left(b\, \frac{\partial}{\partial a} \right) \phi(A) = \sum_{j=1}^{\infty} (\mathrm{Ad}\, A)^{j-1}(B)\, \frac{\phi^{(j)}(A)}{j!}, \tag{58}$$

where $\phi(A) = \sum_{i=0}^{\infty} \nu_i\, A^i$ and

$$\phi^{(j)}(A) = \sum_{i=j}^{\infty} \nu_i\, i(i-1)\cdots(i-j+1)\, A^{i-j}, \quad j \geq 1.$$

We explicitly remark that Baker proves it when $b = Ac$, and then he asserts (without proof) that the same holds true for any arbitrary b.

With these lemmas at hand, Baker (1905, §3) proves his "Exponential Theorem." The proof is the following one. Let us apply (58) when $\phi(A) = e^A = \sum_{i=0}^{\infty} A^i/i!$. In this case, $\phi^{(j)}(A) = e^A$ for every $j \geq 0$ so that (58) gives

$$\left(b\,\frac{\partial}{\partial a}\right) e^A = f(\mathrm{Ad}\,A)(B)\,e^A, \tag{59}$$

where f denotes the formal power series $f(z) := \sum_{j=1}^{\infty} z^{j-1}/j! = (e^z - 1)/z$. This is the very notable identity (analogous to the formula for the differential of the exponential) given by

$$\left(b\,\frac{\partial}{\partial a}\right) e^A = \frac{e^{\mathrm{Ad}\,A} - 1}{\mathrm{Ad}\,A}(B)\,e^A.$$

With this same f, Baker now makes the choice

$$b := \left(1 - \frac{A}{2} + \sum_{j=1}^{\infty} \frac{\varpi_j}{(2j)!} A^{2j}\right) a' = \left(\frac{1}{f}(A)\right) a', \tag{60}$$

where a' is any base and the ϖ_j are the Bernoulli numbers, according to Baker's notation.[36] This implies $a' = f(A)b$ so that, by passing to the corresponding derivatives (recall that the derivative of $A^j b$ is $(\mathrm{Ad}\,A)^j(B)$ by the very notation for the bases),

$$A' = f(\mathrm{Ad}\,A)(B). \tag{61}$$

Gathering (59) and (61), we infer $\left(b\,\frac{\partial}{\partial a}\right) e^A = A'\,e^A$, and inductively

$$\left(b\,\frac{\partial}{\partial a}\right)^i e^A = (A')^i\,e^A, \quad \forall\, i \geq 0. \tag{62}$$

Next Baker puts

$$a'' := \sum_{i=0}^{\infty} \frac{1}{i!} \left(b\,\frac{\partial}{\partial a}\right)^i a \quad \text{and} \quad A'' := \epsilon a''. \tag{63}$$

36 Baker's definition of the Bernoulli numbers is the following one:

$$\frac{z}{e^z - 1} = 1 - \frac{z}{2} + \sum_{j=1}^{\infty} \frac{\varpi_j}{(2j)!} z^{2j}.$$

This agrees with our compact notation, so that $\varpi_j = \mathbf{B}_j$ for every j.

Since Baker formerly proved that ϵ commutes with any power of $b\,\frac{\partial}{\partial a}$ (see Baker 1905, p. 32), we derive from (63)

$$A'' = \sum_{i=0}^{\infty} \frac{1}{i!} \left(b\,\frac{\partial}{\partial a} \right)^{i} A. \tag{64}$$

If we now apply (57) with $t = 1$ and $\mu_j = 1/j!$ (together with the latter representation of A''), then we immediately get

$$e^{A''} = \sum_{i=0}^{\infty} \frac{1}{i!} \left(b\,\frac{\partial}{\partial a} \right)^{i} e^{A} \overset{(62)}{=} e^{A'} e^{A}, \quad \text{whence } e^{A''} = e^{A'} e^{A}. \tag{65}$$

Thus, *the Exponential Theorem is proved once* A'' *in* (64) *is known to be a Lie series in* A, A'. But this is true since A'' is the derivative of an infinite sum of bases, namely a'', and the relation "base $\leftrightarrow$ derivative" has been defined only between Lie elements. Gathering together (60), (64), and (65), Baker has proved the following remarkable formula

$$\begin{aligned}
&e^{A'} e^{A} = e^{A''}, \quad \text{where } A'' = \sum_{i=0}^{\infty} \frac{1}{i!} \left(B\,\frac{\partial}{\partial a} \right)^{i} A, \\
&\text{with} \quad B = A' - \frac{1}{2}[A, A'] + \sum_{j=1}^{\infty} \frac{\varpi_j}{(2j)!} (\mathrm{Ad}\, A)^{2j}(A'), \\
&\text{and where the } \varpi_j \text{ are defined by} \quad \frac{z}{e^z - 1} = 1 - \frac{z}{2} + \sum_{j=1}^{\infty} \frac{\varpi_j}{(2j)!} z^{2j}.
\end{aligned} \tag{66}$$

In confirmation of the computational power of his calculus on bases/derivatives, Baker obtains (in a few lines) his fourth-order expansion (52), starting from the representation of a'' in (63). [Later, but with more elaborate calculations, he will also compute the sixth-order expansion; see Baker (1905, p. 42).] Finally, the last two sections of the paper are devoted to an application of the Exponential Theorem to transformation groups, and in particular to a proof of the Second Fundamental Theorem. Actually, Baker does not add too much to what Pascal (and Campbell) had already proved, nor does he consider in any case the problem of the convergence of the series he obtained.

Though it is full of suggestions for the later history of the CBHD Theorem (indeed Reutenauer 1993 has provided a proof of this theorem which fully formalizes Baker's computations), in reading Baker's proof one inevitably has the feeling that the really important ideas are lost somewhere in the tricky formalism. Hausdorff's argument, which we now turn to analyze, will fill all the gaps.

8 Hausdorff's contribution

Hausdorff[37] devoted a single paper (Hausdorff 1906) to what he defined, in the title of the paper, "the symbolic Exponential Formula in group theory." Despite the mention of the theory of transformation groups, his point of view is made clear from the *incipit* of the paper: "a considerable part of the group theory is completely independent of the actual meaning of the symbols involved" [those of the infinitesimal transformations]. For instance, not even the Jacobi identity depends on the fact that one is dealing with infinitesimal transformations, but only on the form of the bracket $[X, Y] = XY - YX$.[38] Therefore, he proposes to study the function $z = z(x, y)$ defined by the formal identity $e^x e^y = e^z$, a problem of "symbolic analysis," as he defines it.

In the foreword of his paper, Hausdorff provides a brief review of the work of his predecessors. We already mentioned his point of view on Schur, Poincaré, and Pascal. Let us examine what he says about Campbell and Baker. Campbell is acknowledged as the first who "attempted" to give a proof of the Second Theorem of Lie with the aid of a symbolic Exponential Theorem, providing the summands *of degree one* with respect to y in the expansion of $z(x, y)$ (in Hausdorff's subsequent notation, $\chi(y, x)$). Hausdorff's opinion is that Campbell's prolongation of the expansion of the whole z (besides the summands in $\chi(y, x)$) "is based on a passage to the limit not completely clear nor simple."

Even if more than one year elapsed between the two papers,[39] Hausdorff does not mention Baker's paper (1905), and the only reference is to Baker's paper (1901) as containing "some stimulating ideas on the use of matrices in group theory." There is evidence that Hausdorff actually did not know Baker's 1905 paper at the time he published his 1906 manuscript.[40] We will point out every intersection of the two papers in due course.

[37] Felix Hausdorff; Breslau (Prussia; now Wrocław, Poland) 1868—Bonn (Germany), 1942.

[38] Hausdorff used the notation $(X, Y) = XY - YX$ for alternants; consistent with the rest of the paper, we use the usual notation $[X, Y]$.

[39] Baker's front page contains the dates "Received January 8, 1905—Read January 12, 1905," whereas Hausdorff's last page contains the wording "ready for printing February 25, 1906" (literally: *Druckfertig erklärt 25.II.1906.*).

[40] In reviewing Baker's paper (1905), Engel says that "Baker's work, which had escaped Hausdorff, contains a very essential part" [literally, *einen sehr wesentlichen Teil*] "of the results found by Hausdorff." According to Scharlau (*Kommentar*, in Hausdorff 2001, p. 464), it was Engel who informed Hausdorff of Baker's paper. In a letter dated March 9, 1907, Hausdorff wrote back to him: "Following your creditable advice, I have promptly looked at the work by H.F. Baker and sadly noticed that its content really overlaps my work in large part. Evidently the problem has been hanging in the air, since in the last years it has been picked up and solved, with more or less skillfulness, by Campbell, Pascal, Poincaré, Baker; but it is especially Baker's last work—the only one that remained quite unknown to me—that has the most contact with mine. By the way, if I had known it then, I would have published my work nonetheless, albeit with somewhat muted joy, for if compared to all the others, including Baker, it was me who had found the simplest and most transparent proof." [Engel Legacy, Mathematical Institute of Giessen University.] Scharlau furthermore notes that, when Hausdorff was writing to Engel, he had already finished the investigation of the Exponential Formula, for Hausdorff finishes the letter with the words: "Recently, I have been studying set theory only, [...]."

Sections 1 and 2 of Hausdorff (1906) are devoted to describing the necessary algebraic structures and the main operation involved $u\frac{\partial}{\partial x}$, the same as Baker used. If we use modern parlance, Hausdorff in fact introduces the following structures[41]:

L_0 this is the associative algebra (over $\mathbb{R}$ or $\mathbb{C}$) of the polynomials P in a finite set of non-commuting symbols $x, y, z, u, \ldots$; the "dimension" of P is the *least* of the degrees of its monomials.

L this is the associative algebra of the formal power series associated to L_0; any infinite sum is allowed, provided it involves summands with *different* (hence increasing) dimensions.

K_0 this is the Lie subalgebra of L_0 of Lie polynomials in the basis symbols $x, y, z, u, \ldots$

K this is the Lie subalgebra of L of Lie series associated to K_0.

Then Hausdorff considers the important operation on L acting as follows: if $F \in L$ is momentarily thought of as a function of the basis symbol x, and u is a new symbol, we have the "Taylor expansion":

$$F(x+u) = F(x) + \left(u\frac{\partial}{\partial x}\right)F(x) + \frac{1}{2!}\left(u\frac{\partial}{\partial x}\right)^2 F(x) + \frac{1}{3!}\left(u\frac{\partial}{\partial x}\right)^3 F(x) + \cdots, \tag{67}$$

where $\left(u\frac{\partial}{\partial x}\right)^n F(x)$ is the sum of the summands of $F(x+u)$ containing u precisely n times, or equivalently it denotes the action on $F(x)$ of the n-fold composition of the operator $u\frac{\partial}{\partial x}$ with itself, where $u\frac{\partial}{\partial x}$ is the derivation of L mapping x to u and leaving all the other basis symbols unchanged. Even if Baker never wrote down Hausdorff's distinguished expansion (67), the operator $u\frac{\partial}{\partial x}$ is precisely Baker's "substitutional operation" introduced (within the study of the symbolic Exponential Theorem) one year before (see §2 of Baker 1905, p. 31).

For example, if $F = F_0^x + F_1^x + F_2^x + \cdots$, where F_n^x contains x precisely n times, one has

$$\left(x\frac{\partial}{\partial x}\right)F(x) = F_1^x + 2\,F_2^x + \cdots + n\,F_n^x + \cdots. \tag{68}$$

An expansion as in (67) holds true even if u is any element of L, provided that the actual value of u (possibly depending on x) is substituted in the right-hand side of (67) *after* the symbolic operations $\left(u\frac{\partial}{\partial x}\right)^n$ are carried out. From (67) and the commutativity of $u\frac{\partial}{\partial x}$ and $v\frac{\partial}{\partial y}$, one also obtains

$$F(x+u, y+v) = F(x,y) + \sum_{n=1}^{\infty}\frac{1}{n!}\left(u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}\right)^n F(x,y). \tag{69}$$

[41] Lacking of a notion of associative or Lie algebra, any of his structure is generically defined as "Körper."

In order to preserve many of Hausdorff's elegant formulas, we shall use the following notation for *left-nested* iterated brackets:

$$[xy] = [x, y], \quad [xyz] = [[x, y], z], \quad [xyzu] = [[[x, y], z], u], \quad \dots \tag{70}$$

It is then proved in detail that left-nested brackets of basis elements span all K_0, whereas series of such brackets span K. Also, if $F(x)$ is a Lie series in $x, y, z, \dots$, then $\left(u \frac{\partial}{\partial x}\right) F(x)$ is a Lie series in $u, x, y, z, \dots$ (in modern parlance, an algebra derivation is also a Lie algebra derivation).

With only these few prerequisites, Sect. 3 is devoted to the proof of Hausdorff's main result (Hausdorff 1906, Proposition B, p. 29), which is the announced symbolic Exponential Theorem: *The function z of x, y defined by $e^x e^y = e^z$ can be represented as an infinite series whose summands are obtained from x and y by bracket operations, times a numerical factor.* Let us analyze Hausdorff's argument and compare it with the proofs of his predecessors.

We are allowed to say that $z = z(x, y)$ is a function of x, y for in the algebra L the logarithm makes sense (see also Baker's (51)), so that

$$z = t - \frac{1}{2} t^2 + \frac{1}{3} t^3 - \frac{1}{4} t^4 + \cdots, \quad \text{where } t = e^x e^y - 1.$$

We aim to prove that the above z actually belongs to K, not only to L.

Since $\left(u \frac{\partial}{\partial x}\right) e^x = \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{i=0}^{n-1} x^i u x^{n-1-i}$, the substitution $u = [w, x]$ generates a telescopic sum, so that

$$\left([w, x] \frac{\partial}{\partial x}\right) e^x = w e^x - e^x w. \tag{71}$$

Furthermore Hausdorff provides two other well-known formulas of non-commutative algebra[42] (see also the notation in (70)):

$$[w x^n] = \sum_{i=0}^{n} (-1)^i \binom{n}{i} x^i w x^{n-i}, \tag{72}$$

$$e^{-x} w e^x = \sum_{n=0}^{\infty} \frac{1}{n!} [w x^n], \tag{73}$$

where (73) follows easily from (72), by a computation similar to Campbell's (17). As a consequence, we obtain

$$\left([w, x] \frac{\partial}{\partial x}\right) e^x \overset{(71)}{=} e^x (e^{-x} w e^x - w) \overset{(73)}{=} e^x \sum_{n=1}^{\infty} \frac{1}{n!} [w x^n].$$

[42] The formulas (72) and (73) also appear in Campbell (1897a, pp. 387, 386) and in Baker (1905, pp. 34, 38).

An analogous formula (together with the inner factor $(-1)^{n-1}$) with e^x as a right factor holds true. This gives the following results:

If u is of the form $[w, x]$ for some $w \in L$, we have

$$\left(u\,\tfrac{\partial}{\partial x}\right) e^x = e^x\,\varphi(u,x) \quad \textit{where} \quad \varphi(u,x) = \sum_{n=1}^{\infty} \frac{1}{n!}\,[u x^{n-1}], \tag{74}$$

$$\left(u\,\tfrac{\partial}{\partial x}\right) e^x = \psi(u,x)\, e^x \quad \textit{where} \quad \psi(u,x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n!}\,[u x^{n-1}]. \tag{75}$$

We explicitly remark that identity (75) was already discovered by Baker (see (59)). Note that the above identities also hold true when $u = x$, for $\varphi(x,x) = \psi(x,x) = x$ and, by (68), $(x\frac{\partial}{\partial x})e^x = \sum_{n=1}^{\infty} \frac{x^n}{(n-1)!}$.

If we introduce the functions

$$h(z) = \frac{1-e^{-z}}{z}, \qquad g(z) = \frac{1}{h(z)},$$

then φ and ψ can be rewritten as

$$\varphi(u,x) = h(\operatorname{Ad} x)(u), \qquad \psi(u,x) = h(-\operatorname{Ad} x)(u). \tag{76}$$

Furthermore, from (76), we get the inversion formulas:

$$\begin{aligned} p = \varphi(u,x) \;&\Leftrightarrow\; u = \chi(p,x), \\ q = \psi(u,x) \;&\Leftrightarrow\; u = \omega(q,x), \end{aligned} \tag{77}$$

where

$$\chi(p,x) := p - \frac{1}{2}\,[p\,x] + \sum_{n=1}^{\infty} \frac{(-1)^{n-1} B_n}{(2n)!}\,[p\,x^{2n}], \quad \omega(q,x) := \chi(q,-x), \tag{78}$$

and we see how the Bernoulli numbers step in.[43] Note that

$$\chi(p,x) = g(\operatorname{Ad} x)(p), \qquad \omega(q,x) = g(-\operatorname{Ad} x)(q).$$

[43] Indeed we have

$$g(z) = \frac{z}{1-e^{-z}} = 1 + \frac{z}{2} + \sum_{n=1}^{\infty} \frac{(-1)^{n-1} B_n}{(2n)!}\, z^{2n},$$

where, according to Hausdorff's notation, the following definition of Bernoulli numbers B_n holds: $\frac{z}{e^z-1} = 1 - \frac{z}{2} + \sum_{n=1}^{\infty} \frac{(-1)^{n-1} B_n}{(2n)!}\, z^{2n}$. With our notation, it is $B_n = (-1)^{n-1}\mathbf{B}_{2n}$.

Now, identities (74) and (75) and the expansion (67) give the following important formulas (observe that $\varphi(\cdot, x)$ and $\psi(\cdot, x)$ are linear)

$$e^{x+\alpha u} = e^x \left(1 + \alpha\,\varphi(u,x) + \mathcal{O}(\alpha^2)\right), \tag{79}$$
$$e^{x+\alpha u} = \left(1 + \alpha\,\psi(u,x) + \mathcal{O}(\alpha^2)\right) e^x, \tag{80}$$

holding true for every u of the form $[wx]$ (the choice $w = 1$ is also allowed) and every scalar α.[44] We point out that a proof of (79) (with no hypothesis on u) is contained in Poincaré (1900, pp. 244, 245) (see indeed the computation in (40)), but Hausdorff does not mention it, though he knew Poincaré's paper very well.

At this point, Hausdorff's argument becomes less than completely transparent: he states that it is possible to leave z unchanged in $e^x e^y = e^z$ by adding αu to x and by accordingly adding to y a certain quantity $-\alpha v + \mathcal{O}(\alpha^2)$, so that identity

$$e^{x+\alpha u} e^{y-\alpha v+\mathcal{O}(\alpha^2)} = e^z$$

also holds true, *with z unaltered*. We prefer to modify Hausdorff's non-obvious argument in the following way (similar to what is done in Yosida's paper (1937)): let u, v be any pair of elements of L satisfying

$$\varphi(u,x) = \psi(v,y), \tag{81}$$

and let $z(\alpha)$ be defined by

$$e^{z(\alpha)} := e^{x+\alpha\,u} e^{y-\alpha\,v}, \qquad \alpha \in \mathbb{R}.$$

For example, thanks to the inversion formulas (77), the choices

$$(u = x,\ v = \omega(x,y)) \quad \text{or} \quad (v = y,\ u = \chi(y,x)) \tag{82}$$

do satisfy (81). Suppose further that u and v are series of summands of type $[wx]$, so that (79) and (80) apply.[45] (Note that the u's and v's in (82) are of this form.) We thus have the following computation:

$$\begin{aligned} e^{z(\alpha)} &= e^x \left(1 + \alpha\,\varphi(u,x) + \mathcal{O}(\alpha^2)\right)\left(1 - \alpha\psi(v,y) + \mathcal{O}(\alpha^2)\right) e^y \\ &= e^x \left(1 + \alpha(\varphi(u,x) - \psi(v,y)) + \mathcal{O}(\alpha^2)\right) e^y \\ &\overset{(81)}{=} e^x (1 + \mathcal{O}(\alpha^2)) e^y = e^x e^y (1 + \mathcal{O}(\alpha^2)). \end{aligned}$$

[44] Hausdorff never again mentions the restriction on u to be of the form $u = [wx]$. Ipso facto, he will apply (79) and (80) when u is a series of summands of such form, so that this lack of mention is easily understood. However, it is a singular analogy that a similar fact occurred in Baker (1905, p. 34), who proved an analogue of (75) with a restrictive hypothesis similar to Hausdorff's $u = [wx]$, and then Baker stated, without proof, that this restriction could be removed.

[45] Hausdorff fails to mention this hypothesis: this leads us to suppose that he retains (79) and (80) proved with no restriction on u.

From the above expansion, it is easily derived that $\dot{z}(0) = 0$. On the other hand, by applying the expansion (69) to $F(x, y) := \log(e^x e^y)$, we get

$$z(\alpha) = F(x + \alpha\, u, y - \alpha\, v) = z(x, y) + \alpha \left(u \frac{\partial}{\partial x} - v \frac{\partial}{\partial y} \right) z(x, y) + \mathcal{O}(\alpha^2).$$

Hence $\dot{z}(0) = 0$ ensures that, for any u, v satisfying (81), $z(x, y) = e^x e^y$ satisfies the following *PDE* (see Eq. (19) of Hausdorff 1906, p. 28):

$$\left(u \frac{\partial}{\partial x} \right) z = \left(v \frac{\partial}{\partial x} \right) z. \tag{83}$$

We are thus allowed to make, for instance, the choices in (82) which, respectively, give

$$\left(x \tfrac{\partial}{\partial x} \right) z = \left(\omega(x, y) \tfrac{\partial}{\partial y} \right) z, \tag{84}$$

$$\left(y \tfrac{\partial}{\partial y} \right) z = \left(\chi(y, x) \tfrac{\partial}{\partial x} \right) z, \tag{85}$$

and each of these choices suffices to determine z. Indeed, by writing $z = z_0^x + z_1^x + \cdots$ (where the summands in z have been ordered with the increasing powers of x) and by using (68), Hausdorff derives from (84) the following remarkable formula (see Eqs. (22) and (23) of Hausdorff 1906, p. 29):

$$z_n^x = \frac{1}{n!} \left(\omega(x, y) \frac{\partial}{\partial y} \right)^n y, \quad n \geq 0. \tag{86}$$

Here we have used the fact that the application of the operator $\omega(x, y) \frac{\partial}{\partial y}$ increments the degree in x by one unit. Analogously, from (85) one gets

$$z_n^y = \frac{1}{n!} \left(\chi(y, x) \frac{\partial}{\partial x} \right)^n x, \quad n \geq 0. \tag{87}$$

Since x, y and $\omega(x, y)$, $\chi(y, x)$ are all Lie series (see (78)), this proves that any of z_n^x, z_n^y is a Lie polynomial, and the Exponential Theorem is proved.

From (86) and the definition of $\omega(x, y)$, it follows that Hausdorff has proved the following result (note that $[x\, y^{2n}] = (\mathrm{Ad}\, y)^{2n}(x)$):

$$\begin{aligned} &e^x e^y = e^z, \quad \text{where } z = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\omega(x, y) \frac{\partial}{\partial y} \right)^n y, \\ &\text{with } \ \omega(x, y) = x + \frac{1}{2} [x, y] + \sum_{n=1}^{\infty} \frac{(-1)^{n-1} B_n}{(2n)!} (\mathrm{Ad}\, y)^{2n}(x), \\ &\text{and where the } B_n \text{ are defined by } \ \frac{z}{e^z - 1} = 1 - \frac{z}{2} + \sum_{n=1}^{\infty} \frac{(-1)^{n-1} B_n}{(2n)!} z^{2n}. \end{aligned} \tag{88}$$

We explicitly remark that *this is exactly Baker's formula* (66), proved one year earlier.

Despite the fact that Hausdorff was anticipated by Baker, his argument, free from the intricate formalism in Baker (1905), can easily be considered as the first totally transparent proof of the Exponential Theorem; furthermore, it has the merit of uniting contributions from his predecessors (some algebraic computations from Pascal and Campbell; Poincaré's technique of deriving ODEs for $z(\alpha)$; the use of Baker's substitutional operator) in a most effective way.

In order to show the computational effectiveness of his recursion formulas (86) and (87), Hausdorff provides, with a single page of computations, the expansion of z up to order 5 (see the left-nested notation (70)):

$$\begin{aligned} z = x + y &+ \frac{1}{2}[xy] + \frac{1}{12}[xyy] + \frac{1}{12}[yxx] + \frac{1}{24}[yxxy] \\ &- \frac{1}{720}[xyyyy] - \frac{1}{720}[yxxxx] + \frac{1}{360}[xyyyx] + \frac{1}{360}[yxxxy] \\ &- \frac{1}{120}[xyyxy] - \frac{1}{120}[yxxyx] + \cdots \end{aligned}$$

Furthermore, Hausdorff provides a new recursion formula, allowing one to obtain the *homogeneous* summands of z, ordered with respect to the *joint* degree in x, y. His argument is based on his previous techniques: this time, let $z(\alpha)$ be defined by $e^{z(\alpha)} = e^{x+\alpha x}e^y$, so that $z := z(0) = e^x e^y$. By Taylor expansion (67), one has $z(\alpha) = z + \alpha \left(x \frac{\partial}{\partial x}\right) z + \mathscr{O}(\alpha^2)$. Hence, by (80),

$$e^{z(\alpha)} = \left(1 + \alpha\psi\left(x\frac{\partial}{\partial x}z, z\right) + \mathscr{O}(\alpha^2)\right)e^z.$$

On the other hand, again by (80) applied to $e^{x+\alpha x}$, we infer

$$e^{z(\alpha)} = e^{x+\alpha x}e^y = \left(1 + \alpha\,\psi(x,x) + \mathscr{O}(\alpha^2)\right)e^x e^y = \left(1 + \alpha\, x + \mathscr{O}(\alpha^2)\right)e^z.$$

By comparing the above two expansions of $e^{z(\alpha)}$, we derive $x = \psi\left(x\frac{\partial}{\partial x}z, z\right)$, so that, by the inversion formula (77) for ψ, we get

$$\left(x\frac{\partial}{\partial x}\right)z = \omega(x,z). \tag{89}$$

A dual argument for $e^x e^{y+\alpha y}$ yields another "PDE"

$$\left(y\frac{\partial}{\partial y}\right)z = \chi(y,z). \tag{90}$$

Gathering together (89) and (90), we obtain

$$\begin{aligned}
&\left(x\,\frac{\partial}{\partial x}\right)z+\left(y\,\frac{\partial}{\partial y}\right)z=\omega(x,z)+\chi(y,z)\\
&\overset{(78)}{=} x+\tfrac{1}{2}\,[xz]+\sum_{n=1}^{\infty}\tfrac{(-1)^{n-1}B_n}{(2n)!}\,[x\,z^{2n}]+y-\tfrac{1}{2}\,[yz]+\sum_{n=1}^{\infty}\tfrac{(-1)^{n-1}B_n}{(2n)!}\,[y\,z^{2n}]\\
&=x+y-\frac{1}{2}\,[x+y,z]+\sum_{n=1}^{\infty}\frac{(-1)^{n-1}B_n}{(2n)!}\,(\mathrm{Ad}\,z)^{2n}(x+y)\\
&=[x,z]+\chi(x+y,z).
\end{aligned}$$

This gives the remarkable formula (see Eq. (29) of Hausdorff 1906, p. 31):

$$\left(x\,\frac{\partial}{\partial x}\right)z+\left(y\,\frac{\partial}{\partial y}\right)z=[x,z]+\chi(x+y,z). \tag{91}$$

Inserting into (91) the expansion $z=z_1^{x,y}+z_2^{x,y}+\cdots$, where $z_n^{x,y}$ has joint degree n in x,y, we obtain a recursion formula for the summands $z_n^{x,y}$ which exhibits in a very limpid form their Lie-polynomial nature. This formula will return in modern proofs of the Campbell–Baker–Hausdorff Theorem: (see Djoković 1975; Varadarajan 1984) and, as shown in Varadarajan (1984), it can be profitably used for convergence matters.

From Sect. 4 onward, Hausdorff turns his attention to the applications of the symbolic Exponential Theorem to groups of transformations. Since he blamed his predecessors (in particular Pascal, who indeed got very close to proving a symbolic theorem) for omitting convergence matters, Hausdorff's main concern is to prove the convergence of at least one of his three series-expansions of $z(x,y)$, for x and y near the origin.

To this end, let $t_1,\dots,t_r$ be a basis of a finite dimensional set of infinitesimal transformations,[46] with structure constants given by $[t_\rho,t_\sigma]=\sum_{\lambda=1}^{r}c_{\rho\sigma\lambda}t_\lambda$. Let $x=\sum_\rho \xi_\rho t_\rho$, $y=\sum_\rho \eta_\rho t_\rho$ and suppose that the series $z(x,y)=\sum_{n=0}^{\infty}z_n^y$ defined by (87) converges to $z=\sum_\rho \zeta_\rho t_\rho$. It must be remarked that, in its turn, any z_n^y (with $n\geq 1$) is given by a series, convergence of which must be proved first. Hausdorff begins with the study of $z_1^y=\chi(y,x)$. Suppose this converges to $u=\sum_\rho \vartheta_\rho t_\rho$. By definition of $u=\chi(y,x)$ and by inversion $y=\varphi(u,x)$ (recalling (77)), we see that u depends linearly on y and so y depends linearly on u. Passing to the associated coordinates with respect to the basis $\{t_1,\dots,t_r\}$, we infer the existence of two matrices:

$$A(\xi)=\big(\alpha_{\rho\sigma}(\xi)\big),\qquad \Psi(\xi)=A(\xi)^{-1}=\big(\psi_{\rho\sigma}(\xi)\big),$$

46 As can be seen, Hausdorff's argument works very well for any finite dimensional Lie algebra.

such that $\vartheta_\sigma = \sum_\rho \alpha_{\rho\sigma}(\xi)\,\eta_\rho$ and $\eta_\rho = \sum_\sigma \psi_{\sigma\rho}(\xi)\,\vartheta_\sigma$, so that the identity $u = \chi(y,x)$ gives us the compact relation:

$$\sum_\sigma \alpha_{\rho\sigma}(\xi)\,t_\sigma = \chi\left(t_\rho, \sum_\lambda \xi_\lambda t_\lambda\right), \quad \rho = 1,\dots,r. \tag{92}$$

This allows Hausdorff to obtain an explicit series for the functions $\alpha_{\rho\sigma}(\xi)$ and, consequently, for the functions ϑ_ρ. To this end, let us introduce the structure matrix

$$\Xi(\xi) := (\xi_{\rho\sigma}), \quad \text{where } \xi_{\rho\sigma} := \sum_{\lambda=1}^{r} c_{\rho\lambda\sigma}\xi_\lambda.$$

(This is nothing but the transpose matrix of the matrix representing the right-adjoint map $Y \mapsto [Y, \sum_\lambda \xi_\lambda t_\lambda]$.) By means of the matrix Ξ, we can write down explicitly the left-nested brackets $[t_\rho x^n]$. This allows us to recognize that (92) is equivalent to

$$\alpha_{\rho\sigma}(\xi) = \delta_{\rho\sigma} - \frac{1}{2}\xi_{\rho\sigma} + \frac{B_1}{2!}\sum_\lambda \xi_{\rho\lambda}\xi_{\lambda\sigma} - \frac{B_2}{4!}\sum_{\lambda\mu\nu}\xi_{\rho\lambda}\xi_{\lambda\mu}\xi_{\mu\nu}\xi_{\nu\sigma} + \cdots$$

Hausdorff recognizes that this series had already been written by Schur (1890a,b), but he adds that it can be "elegantly rewritten" as follows (see Eq. (11) of Hausdorff 1906, p. 34):

$$\begin{aligned} &A = 1 - \frac{1}{2}\Xi + \frac{B_1}{2!}\Xi^2 - \frac{B_2}{4!}\Xi^4 + \cdots, \\ &\text{i.e.,}\quad A(\xi) = f(\Xi(\xi)), \qquad \text{where } f(z) = \frac{z}{e^z - 1}. \end{aligned} \tag{93}$$

It is now very simple to deduce from (93) a domain of convergence for the series expressing the functions $\alpha_{\rho\sigma}(\xi)$: if M is an upper bound for all the functions $|\xi_{\rho\sigma}|$, it suffices to have $rM < 2\pi$, since the complex function $f(z)$ is holomorphic in the disk about 0 of radius 2π. This produces a domain of convergence for $\vartheta_\sigma = \sum_\rho \alpha_{\rho\sigma}(\xi)\,\eta_\rho$ and hence of $u = \sum_\rho \vartheta_\rho t_\rho$, which is the first summand $\chi(y,x)$ in the expansion for $z(x,y)$.

Before passing to the convergence of the other summands and of the whole series of these summands, Hausdorff compares his matrix-calculus to Baker's (1901). Let us introduce the matrix

$$\Theta(\xi) := (\theta_{\rho\sigma}(\xi)) \quad \text{such that } e^{-x}t_\rho e^x = \sum_\sigma \theta_{\rho\sigma}t_\sigma.$$

By means of (73), we derive the matrix identity

$$\Theta(\xi) = e^{\Xi(\xi)}.$$

Now, from the Exponential Theorem $e^x e^y = e^z$, we see that the following chain of equality holds

$$\sum_\tau \theta_{\rho\tau}(\zeta)t_\tau = e^{-z}t_\rho e^z = e^{-y}(e^{-x}t_\rho e^x)e^y = e^{-y}\left(\sum_\sigma \theta_{\rho\sigma}(\xi)t_\sigma\right)e^y$$
$$= \sum_\sigma \theta_{\rho\sigma}(\xi)\sum_\tau \theta_{\sigma\tau}(\eta)t_\tau = \sum_\tau\left(\sum_\sigma \theta_{\rho\sigma}(\xi)\theta_{\sigma\tau}(\eta)\right)t_\tau.$$

By comparing the coefficients of the t_τ's, we get the identities

$$\Theta(\zeta) = \Theta(\xi)\,\Theta(\eta), \qquad e^{\Xi(\xi)}\,e^{\Xi(\eta)} = e^{\Xi(\zeta)}. \tag{94}$$

According to Hausdorff, presumably unaware of Baker's paper (1905), this is "Baker's Exponential Theorem," a corollary of his symbolic theorem.

Hausdorff returns to the convergence matter in Sect. 6. This is his argument. He aims to prove the convergence (near the origin) of

$$z = x + \chi(y,x) + \frac{1}{2!}\left(\chi(y,x)\frac{\partial}{\partial x}\right)\chi(y,x) + \frac{1}{3!}\left(\chi(y,x)\frac{\partial}{\partial x}\right)^2\chi(y,x) + \cdots. \tag{95}$$

Let us set $u := \chi(y,x)$; it suffices to discover what the operator $u\,\frac{\partial}{\partial x}$ looks like in coordinates with respect to $t_1,\dots,t_r$. Recalling the notation $u = \sum_\rho \vartheta_\rho t_\rho$ and setting also $F(x) = \sum_\rho f_\rho(\xi)\,t_\rho$, we have the usual Taylor expansion

$$F(x+u) = \sum_\rho f_\rho(\xi+\vartheta)\,t_\rho = \sum_\rho\left(f_\rho(\xi) + \sum_\sigma \vartheta_\sigma\frac{\partial f_\rho}{\partial \xi_\sigma}(\xi) + \mathscr{O}(|\vartheta|^2)\right)t_\rho,$$

which, compared to the expansion $F(x+u) = F(x) + u\,\frac{\partial}{\partial x}F(x) + \mathscr{O}(u^2)$ in (67), shows that the operator $u\,\frac{\partial}{\partial x}$ has the same meaning as the infinitesimal transformation $\Lambda = \sum_{\sigma=1}^r \vartheta_\sigma(\xi,\eta)\,\frac{\partial}{\partial \xi_\sigma}$. Recalling that $\vartheta_\sigma = \sum_\rho \alpha_{\rho\sigma}(\xi)\,\eta_\rho$, we obtain the equivalent expression for Λ

$$\Lambda = \sum_{\rho,\sigma=1}^r \eta_\rho\,\alpha_{\rho\sigma}(\xi)\,\frac{\partial}{\partial \xi_\sigma}. \tag{96}$$

This proves that the expansion (95) becomes, using coordinates $z = \sum_\rho \zeta_\rho\,t_\rho$,

$$\zeta_\rho = \xi_\rho + \Lambda\xi_\rho + \frac{1}{2!}\Lambda^2\xi_\rho + \frac{1}{3!}\Lambda^3\xi_\rho + \cdots =: e^\Lambda(\xi_\rho). \tag{97}$$

Now, the crucial insight is to observe that this is precisely the expansion of the solution $t \mapsto \zeta(t)$ to the ODE system

$$\begin{cases} \dot{\zeta}_\sigma(t) = \sum_\rho \eta_\rho \, \alpha_{\rho\sigma}(\zeta(t)) \\ \zeta_\sigma(0) = \xi_\sigma \end{cases} \qquad \sigma = 1, \dots, r.$$

Hence, since it has already been proved that the maps $\alpha_{\rho\sigma}$ are analytic in a neighborhood of the origin, *the convergence of the functions ζ_ρ in* (97) *is a consequence of the general theory of ODEs.* Note the similarity with Poincaré's convergence argument, see (38), (39).

Furthermore, Hausdorff suggests an evocative way of interpreting his PDE-like operators: for instance,

$$y \frac{\partial}{\partial y} z = \chi(y, z) \quad \text{is to} \quad z = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\chi(y, x) \frac{\partial}{\partial x} \right)^n x \quad \text{as}$$

$$\dot{\zeta}_\sigma = \sum_\rho \eta_\rho \, \alpha_{\rho\sigma}(\zeta) \quad \text{is to} \quad \zeta = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\sum_{\rho,\sigma} \eta_\rho \, \alpha_{\rho\sigma}(\xi) \frac{\partial}{\partial \xi_\sigma} \right)^n \xi.$$

Finally, the above convergence result furnishes the proof of the *transformation-group Exponential Theorem* (Hausdorff 1906, Proposition C, p. 39): *Given the structure equations $[t_\rho, t_\sigma] = \sum_\lambda c_{\rho\sigma\lambda} t_\lambda$, by means of the Exponential Theorem*

$$e^{\sum \xi_\rho t_\rho} e^{\sum \eta_\rho t_\rho} = e^{\sum \zeta_\rho t_\rho},$$

the functions $\zeta_\rho = \zeta_\rho(\xi, \eta)$ are well defined and analytic in a suitable neighborhood of $\xi = \eta = 0$.

Section 7 is devoted to the derivation of the Second and Third Theorems of Lie by means of the above Proposition C. As a matter of fact, Hausdorff's argument does not add too much to what his predecessors, from Poincaré to Baker, had said about the same topic.

In closing this review about Hausdorff's contribution to the Exponential Theorem, we quote his words (see Hausdorff 1906, p. 44): the symbolic Exponential Theorem "Proposition B is the *nervus probandi* of the fundamental theorems of group theory." An opinion shared, with the same fervor, by Poincaré and Pascal.

9 Dynkin's contribution

After Hausdorff's 1906 paper, more than 40 years elapsed before the problem of providing an explicit representation of the series $\log(e^x e^y)$ was solved. This question was first answered by Dynkin[47] in his 1947 paper. In what follows, we will quote the 2000

[47] Eugene Borisovich Dynkin; Leningrad (Union of Soviet Socialist Republics), 1924.

English translation of the Russian paper (Dynkin 1947), contained in Dynkin (2000, pp. 31–34).

Starting from what Dynkin calls "the theorem of Campbell and Hausdorff," i.e., the result stating that $\log(e^x e^y)$ is a series of Lie polynomials in x and y, formula (12) of Dynkin (2000) provides the following explicit representation (notations will be explained below), later known as *Dynkin's Formula* (for the Campbell–Baker–Hausdorff series)

$$\log(e^x e^y) = \sum \frac{(-1)^{k-1}}{k} \frac{1}{p_1!q_1!p_2!q_2!\cdots p_k!q_k!} \left(x^{p_1} y^{q_1} x^{p_2} y^{q_2} \cdots x^{p_k} y^{q_k}\right)^0, \tag{98}$$

where the sum runs over $k \in \mathbb{N}$ and the non-vanishing couples (p_i, q_i), with $i = 1, \ldots, k$. Most importantly, the map $P \mapsto P^0$, which we now describe, is introduced, where P is any polynomial in a finite set of non-commuting indeterminates.

In order to define the map $P \mapsto P^0$, let $\mathscr{R}$ denote the (free) algebra of the polynomials in the non-commuting indeterminates $x_1, \ldots, x_n$ over a field of characteristic zero. For $P, Q \in \mathscr{R}$, let $[P, Q] = PQ - QP$ denote the usual commutator,[48] and let $\mathscr{R}^0$ be the least Lie subalgebra of $\mathscr{R}$ containing $x_1, \ldots, x_n$. Finally, let us consider the unique linear map from $\mathscr{R}$ to $\mathscr{R}^0$ mapping $P = x_{i_1} x_{i_2} \cdots x_{i_k}$ to P^0, where

$$P^0 = \frac{1}{k} [\cdots [[x_{i_1}, x_{i_2}], x_{i_3}] \cdots x_{i_k}].$$

Then (see Theorem on page 32 of Dynkin 2000) Dynkin proves that

$$P \in \mathscr{R}^0 \textit{ if and only if } P = P^0. \tag{99}$$

This theorem, later also referred to as the Dynkin–Specht–Wever Theorem (see also Specht 1948; Wever 1949), is one of the main characterizations of Lie polynomials.

Dynkin's proof of the above theorem reads as follows. We only have to prove that $P \in \mathscr{R}^0$ implies $P = P^0$, and it is non-restrictive to suppose that P is an ordered Lie-monomial, say $P = [\cdots [x_1, x_2] \cdots x_m]$. By expanding the brackets, we obtain uniquely determined scalars $a_{i_1 \ldots i_m}$ such that

$$P = [\cdots [x_1, x_2] \cdots x_m] = \sum_{I=(i_1,\ldots,i_m)} a_{i_1 \ldots i_m} x_{i_1} \cdots x_{i_m}, \tag{100}$$

where I is any permutation of $\{1, \ldots, m\}$. Hence, by the very definition of P^0,

$$P^0 = \sum_{I=(i_1,\ldots,i_m)} \frac{1}{m} a_{i_1 \ldots i_m} [\cdots [x_{i_1}, x_{i_2}] \cdots x_{i_m}]. \tag{101}$$

[48] Dynkin used the notation $P \circ Q := PQ - QP$; also $x_{i_1} \circ x_{i_2} \circ \cdots \circ x_{i_k}$ denotes the left-nested commutator $(\cdots((x_{i_1} \circ x_{i_2}) \circ x_{i_3}) \circ \cdots \circ x_{i_k})$. We allowed ourselves to use the bracket notation, as in the rest of the paper.

Now, by skew-symmetry and Jacobi identity $[[v, w], u] = [[u, w], v] - [[u, v], w]$, for any fixed $k \in \{1, \dots, m\}$ one has

$$[\cdots[x_1, x_2]\cdots x_m] = \sum_{J=(j_2,\dots,j_m)} c_{kj_2\dots j_m}\,[\cdots[x_k, x_{j_2}]\cdots x_{j_m}], \tag{102}$$

where J is any permutation of $\{1, \dots, m\} \setminus \{k\}$. If we expand the generic summand in the right-hand side of (102), then we get (among all monomials) a unique monomial which starts with x_k, namely $x_k x_{j_2} \cdots x_{j_m}$: thus, comparing (100) and (102), we derive $c_{kj_2\dots j_m} = a_{kj_2\dots j_m}$ for every J as above, so that

$$[\cdots[x_1, x_2]\cdots x_m] = \sum_{J=(j_2,\dots,j_m)} a_{kj_2\dots j_m}\,[\cdots[x_k, x_{j_2}]\cdots x_{j_m}]. \tag{103}$$

Being $k \in \{1, \dots, m\}$ arbitrary, summing up we obtain

$$m\,P = m\,[\cdots[x_1, x_2]\cdots x_m] = \sum_{I=(i_1,i_2,\dots,i_m)} a_{i_1 i_2\dots i_m}\,[\cdots[x_{i_1}, x_{i_2}]\cdots x_{i_m}],$$

where I is as in (100). By comparing this identity to (101), we infer $P = P^0$, and (99) is proved.

With this result at hand, the derivation of the representation (98) is almost trivial. Indeed, the very definitions of log and exp give

$$\log(e^x e^y) = \sum \frac{(-1)^{k-1}}{k} \frac{1}{p_1!q_1!p_2!q_2!\cdots p_k!q_k!}\, x^{p_1}y^{q_1}x^{p_2}y^{q_2}\cdots x^{p_k}y^{q_k},$$

where the sum is as in (98). *If we assume that the Exponential Theorem is satisfied* (i.e., that the above series is indeed a Lie series in x and y), an application of the map $P \mapsto P^0$ (naturally extended to series) leaves unchanged $\log(e^x e^y)$ so that (98) holds true.

Other three fundamental results are contained in Dynkin's 1947 paper (see Dynkin 2000, p. 34):

1. If $\mathbb{K}$ is $\mathbb{R}$ or $\mathbb{C}$ and $\mathfrak{g}$ is any finite dimensional Lie algebra over $\mathbb{K}$, then the series in the right-hand side of (98), say $\Phi^0(x, y)$, converges for every x, y in a neighborhood of $0 \in \mathfrak{g}$. Indeed, *thanks to the very explicit expression of Dynkin's series*, the following direct computation holds: if $\|\cdot\|$ is any norm on $\mathfrak{g}$ compatible with the Lie bracket,[49] then

$$\left\| \left(x^{p_1}y^{q_1}x^{p_2}y^{q_2}\cdots x^{p_k}y^{q_k}\right)^0 \right\| \le \|x\|^{p_1+\cdots+p_k} \cdot \|y\|^{q_1+\cdots+q_k}.$$

[49] This means that $\|[x, y]\| \le \|x\| \cdot \|y\|$ for every $x, y \in \mathfrak{g}$. Note that such a norm always exists, thanks to the continuity and bilinearity of $(x, y) \mapsto [x, y]$ and the finite dimensionality of $\mathfrak{g}$.

Consequently, as for the study of the absolute convergence of the series $\Phi^0(x, y)$, an upper bound is given by

$$\sum_k \frac{(-1)^{k-1}}{k}\left(e^{\|x\|}e^{\|y\|} - 1\right)^k = \log(e^{\|x\|}e^{\|y\|}) = \|x\| + \|y\| < \infty,$$

provided that $\|x\| + \|y\| < \log 2$. As a matter of fact, *this is the first argument in the history of the CBHD Theorem for a direct proof of the convergence problem*, an argument much more natural and direct than those given by Poincaré, by Pascal, and even by Hausdorff (whose convergence-study was the most satisfactory, up to that time). Also, Dynkin's proof works *for any finite dimensional Lie algebra*, hence in particular for the algebra of the infinitesimal transformations of a transformation group, thus comprising all the related results obtained by his predecessors.

2. The same arguments as above can be straightforwardly generalized to the so-called *Banach–Lie algebras*,[50] thus anticipating a newer research field. Dynkin will extensively return to this generalization in his paper (1950).
3. Dynkin's series (98), together with the obvious (local) associativity of the (local) operation $x * y := \Phi^0(x, y)$, allows one to attack the original version of Lie's Third Theorem (a concern for Schur, Poincaré, Pascal, and Hausdorff) in a very simple way: indeed $*$ defines a *local group* on a neighborhood of the origin of every finite dimensional (real or complex) Lie algebra, with prescribed structure constants.

As observed above, Dynkin proves a commutator-formula for $\log(e^x e^y)$ yet assuming its commutator-nature. Two years later, Dynkin (1949) will give another proof of the fact that $\log(e^x e^y)$ is a Lie series, completely independent of the arguments of his predecessors, and mainly based on his theorem (99) and on combinatorial computations. Quoting the words of Bose (1989, p. 2035), with respect to the recursive formulas for $\log(e^x e^y)$ proved by Baker and Hausdorff, "Dynkin radically simplified the problem," by deriving "an effective procedure" for determining the Campbell–Baker–Hausdorff series. Let us overview his new proof.

Let $\mathbb{K}$ be a field of characteristic zero. We set $\Phi(x, y) = \log(e^x e^y)$, and we denote by $P_{p,q}(x, y)$ a sum analogous to that in the right-hand side of (98) extended over the indices $p_1 + \cdots + p_k = p$ and $q_1 + \cdots + q_k = q$. Thus we have

$$\log(e^x e^y) = \sum_{p,q \geq 0} P_{p,q}(x, y), \tag{104}$$

and Dynkin aims to prove that $P_{p,q}(x, y)$ is a Lie polynomial in x, y. More generally, let us consider the algebra of the polynomials in the non-commuting indeterminates $x_0, x_1, \ldots, x_n$. If $t_0, \ldots, t_n \in \mathbb{Q}$, let us set $\Phi(t_0 x_0, \ldots, t_n x_n) := \log(e^{t_0 x_0} \cdots e^{t_n x_n})$.

[50] A Banach–Lie algebra is a Banach space (over $\mathbb{R}$ or $\mathbb{C}$) endowed with a Lie algebra structure such that $A \times A \ni (x, y) \mapsto [x, y] \in A$ is continuous. In this case, if $\|\cdot\|$ is the norm of A, there exists a positive constant M such that $\|[x, y]\| \leq M\,\|x\| \cdot \|y\|$ for every $x, y \in A$, so that the norm $M\|\cdot\|$ is compatible with the Lie bracket of A, and Dynkin's arguments (this time also appealing to the completeness of A) can be directly generalized.

Then we have the series expansion

$$\Phi(t_0 x_0, \dots, t_n x_n) = \sum_{I=(i_0,\dots,i_n)} t_0^{i_0} \cdots t_n^{i_n} \, P_I(x_0, \dots, x_n), \tag{105}$$

where $P_I(x_0, \dots, x_n)$ is a polynomial of degree i_0 in x_0, i_1 in x_1 and so on. If $x_0 = \cdots = x_{p-1} = x$ and $x_p = \cdots = x_n = y$, then one obtains

$$\log\left(e^{(t_0+\cdots+t_{p-1})x} e^{(t_p+\cdots+t_n)y}\right) = \sum_I t_0^{i_0} \cdots t_n^{i_n} \, P_I \left(\underbrace{x, \dots, x}_{p \text{ times}}; \underbrace{y, \dots, y}_{q \text{ times}} \right), \tag{106}$$

where $p + q = n + 1$. On the other hand, recalling (104), the left-hand side of (106) equals

$$\sum_{p,q} (t_0 + \cdots + t_{p-1})^p \, (t_p + \cdots + t_n)^q \, P_{p,q}(x, y) = \sum_{p,q} P_{p,q}(x, y)$$
$$\times \sum_{i_0+\cdots+i_{p-1}=p} \frac{p!}{i_0! \cdots i_{p-1}!} t_0^{i_0} \cdots t_{p-1}^{i_{p-1}} \sum_{i_p+\cdots+i_n=q} \frac{q!}{i_p! \cdots i_n!} t_p^{i_p} \cdots t_n^{i_n}.$$

By equating the summands with all the t's of degree 1 from the above far right-hand sum and from the right-hand side of (106), Dynkin obtains the equality (see Eq. (6) of Dynkin 1949, p. 157):

$$P_{p,q}(x, y) = \frac{1}{p!\,q!} \, P \left(\overbrace{x, \dots, x}^{p \text{ times}}; \overbrace{y, \dots, y}^{q \text{ times}} \right), \quad \text{where } P := P_{(1,\dots,1)}. \tag{107}$$

All the attention is now shifted to the polynomial P. By the definition of Φ, one recognizes that P can be obtained also as the sum of the coefficients of $t_0, t_1, \dots, t_n$ of degree 1 in the expansion of

$$\log\left((1 + t_0 x_0) \cdots (1 + t_n x_n)\right) = -\sum_{\alpha=1}^{\infty} \tfrac{(-1)^\alpha}{\alpha} \left(\sum_{j_0 < \cdots < j_k} t_{j_0} \cdots t_{j_k} \, x_{j_0} \cdots x_{j_k} \right)^{\alpha}. \tag{108}$$

The term $t_0 \cdots t_n \, x_{i_0} \cdots x_{i_n}$ appears in expanding the above series so many times as one can subdivide $(i_0, \dots, i_n)$ in α finite sequences, each monotone increasing, chosen from $\{0, 1, \dots, n\}$. Then Dynkin introduces a useful definition: in the $(n + 1)$-tuple $I = (i_0, \dots, i_n)$ of pairwise distinct integers, a couple of consecutive indices $i_\beta, i_{\beta+1}$ is called *regular* if $i_\beta < i_{\beta+1}$ and *irregular* otherwise. (Now these are also called, respectively, the *rises* and the *falls*.) He denotes by s_I the number of regular couples of I and by t_I the number of irregular ones. He then shows that the coefficient of $t_0 \cdots t_n \, x_{i_0} \cdots x_{i_n}$ in (108) is equal to

$$-\sum_{\alpha=t_I+1}^{n} \frac{(-1)^\alpha}{\alpha} \binom{s_I}{\alpha - t_I - 1} = (-1)^{t_I} \frac{s_I!\,t_I!}{(n+1)!}.$$

This gives the representation (see Eq. (8) in Dynkin 1949, p. 158)

$$P(x_0, \dots, x_n) = \frac{1}{(n+1)!} \sum_I b_I\, x_{i_0} \cdots x_{i_n}, \quad \text{with } b_I := (-1)^{t_I}\, s_I!\, t_I!, \tag{109}$$

where the sum runs over the permutations I of $\{0, 1, \dots, n\}$. The main task is thus to prove that the right-hand side of (109) is a Lie polynomial. This is accomplished in Sect. 2.

To this end, Dynkin improves his former theorem (99). Indeed, he claims that with the same arguments as in Dynkin (1947) the following result holds true: *Let Q be a polynomial in $x_0, \dots, x_n$ of the following special form: $Q = \sum_I a_I\, x_{i_0} \cdots x_{i_n}$, the sum ranging over the permutations of $\{0, \dots, n\}$. Then Q is a Lie polynomial if and only if*

$$Q = \sum_J a_{0J}\, [\cdots [[x_0, x_{j_1}], x_{j_2}] \cdots x_{j_n}],$$

where the sum ranges over the permutations $J = (j_1, \dots, j_n)$ of $\{1, \dots, n\}$.

Since P in (109) has the form of the above Q, this suggests to prove that

$$\sum_I b_I\, x_{i_0} \cdots x_{i_n} = \sum_J b_{0J}\, [\cdots [[x_0, x_{j_1}], x_{j_2}] \cdots x_{j_n}], \tag{110}$$

where the coefficients b_I, b_{0J} are as in (109), and where $I = (i_0, \dots, i_n)$ runs over the permutations of $\{0, 1, \dots, n\}$, whereas $J = (j_1, \dots, i_n)$ runs over the permutations of $\{1, \dots, n\}$. Since the right-hand side of (110) is a Lie polynomial, this will prove, because of (104) and (107), the theorem of Campbell–Baker–Hausdorff.

The proof of (110) is technical and quite laborious, and the reader is directly referred to Dynkin (1949, pp. 159–161). However, some remarks on this proof are due, since it contains the concept of "shuffling." Let us expand the right-hand side of (110) in monomials, say $\sum_I c_I\, x_{i_0} \cdots x_{i_n}$. We have to prove that $c_I = b_I$ for all indices I as above. Let us split I in two disjoint sequences

$$I' = (i'_1, \dots, i'_p), \quad I'' = (i''_1, \dots, i''_q) \qquad (p + q = n + 1),$$

where the i' and i'' are written in the same order as they appear in I. In this case, we say that I is a *shuffle* of I' and I'', and we write (in Dynkin's notation) $I \in I' \cup I''$. The crucial remark, allowing us to transform (110) into a problem of combinatorics, is the following: let x_0 occur in $x_{i_0} \cdots x_{i_n}$ in the position $p + 1$, and let us use the notation:

$$x_{i_0}\cdots x_{i_n} = x_{I'}\,x_0\,x_{I''} =: x_{I'0I''},$$

where $I' = (i_0, \dots, i_{p-1})$ and $I'' = (i_{p+1}, \dots, i_n)$. It is easy to see that, by expanding the higher order commutator $[\cdots[[x_0, x_{j_1}], x_{j_2}]\cdots x_{j_n}]$ in monomials, the summand $x_{I'0I''}$ occurs if and only if $J = (j_1, \dots, j_n)$ is a shuffle of $\widetilde{I'} := (i_{p-1}, \dots, i_1, i_0)$ (the reversed of I') and of I''; in this case, $x_{I'0I''}$ has coefficient $+1$ or -1 according to whether p is even or odd.

This gives the formula

$$c_{I'0I''} = (-1)^p \sum_{J\in\widetilde{I'}\cup I''} b_{0J}.$$

The rest of the proof consists of showing that this last identity is satisfied when the c's in its left-hand side are replaced by the b's introduced in (109).

Summing up all his identities, Dynkin is able to provide yet another Lie representation of $\log(e^x e^y)$, namely (gathering together (104), (107), (109), and (110)):

$$\log(e^x e^y) = \sum_{p,q=0}^{\infty} \frac{1}{p!\,q!} P\Big(\underbrace{x,\dots,x}_{p\text{ times}};\underbrace{y,\dots,y}_{q\text{ times}}\Big), \quad \text{where}$$
$$P(x_0,\dots,x_n) = \frac{1}{(n+1)!}\sum_J (-1)^{t_{0J}}\, s_{0J}!\, t_{0J}!\, [[x_0, x_{j_1}]\cdots x_{j_n}],$$

or, in a more symmetric form,

$$P(x_0,\dots,x_n) = \frac{1}{(n+1)\,(n+1)!}\sum_I (-1)^{t_I}\, s_I!\, t_I!\, [[x_{i_0}, x_{i_1}]\cdots x_{i_n}],$$

where $J=(j_1,\dots,i_n)$ runs over the permutations of $\{1,\dots,n\}$, while $I=(i_0,\dots,i_n)$ runs over the permutations of $\{0,1,\dots,n\}$.

Finally, in his paper (1950) Dynkin studies in great detail the applications of his representation formula for $\log(e^x e^y)$ to normed Lie algebras and to analytic groups. The *incipit* of the paper clearly states that the "one-to-one correspondence between local Lie groups and Lie algebras is the foundation of Lie groups." He then highlights that the algorithms of Campbell and Hausdorff (and also of Baker, we may add, though his papers are not mentioned) are not effective for solving the problem of convergence, making it necessary "to turn to an indirect proof, using Lie's system of differential equations" and making use of ODEs theory: an application of his representation formulas from Dynkin (1947, 1949) "eliminates the need for any complementary *apparatus*."

Also the point of view of Dynkin (1950) is new: the starting points are not the transformation groups, but the Lie algebras. The main result is thus the *construction of a local topological group attached to every Banach–Lie algebra by means of the explicit series of* $\log(e^x e^y)$. The theory of groups and algebras was meanwhile advanced sufficiently to allow it possible to use more general notions and provide broader generalizations (non-Archimedean fields are considered; normed spaces and normed algebras are involved, together with local topological or analytic groups). As for the history of the CBHD Theorem, this paper paved the way for the problem of *improved domains of convergence* for the Dynkin series and, by happenstance, for the study of *other possible representations* of $\log(e^x e^y)$.

To these topics, the modern literature on the CBHD Theorem has devoted a great attention. We refer the reader to the references in Bonfiglioli and Fulci (2012, Sect. 5.6). In particular, for the intertwining of the CBHD Theorem and its "continuous version," the so-called Magnus Expansion, see Blanes et al. (2009) and the comprehensive list of references therein.

10 Hausdorff and Bourbaki's commentaries

Along with the fact that there is still no conventional agreement on the "fatherhood" of the CBHD Theorem (Baker's name is frequently unfairly omitted), the most surprising fact concerning the early investigations is that the contributions by Schur, Poincaré, and Pascal have presently become almost forgotten.[51] In our opinion, there are three reasons for this fact. First, they focused on the setting of Lie's theory of groups of transformations, instead of taking up a purely symbolic approach. Second, the very cold judgement of the contributions of Schur, Poincaré, and Pascal (and also of Campbell and Baker) given by Hausdorff himself and by Bourbaki has certainly played a major rôle. Third, the fact that none of their original papers on the subject was written in English has contributed to their becoming obsolete. (Hausdorff's paper did not share this lot, though.)

First, we begin by considering Hausdorff's and Bourbaki's comments on the works by Schur, Poincaré, and Pascal; then, we express our opinion on these comments. The reader will certainly make his/her own opinion by rediscovering the contributions of these three mathematicians contained in Sects. 3, 5, and 6.

In Hausdorff (1906, p. 20), we find the following comments:

- Poincaré, in handling "Campbell's problem," has *not* proved the Exponential Theorem in its general and *symbolic* form, but instead in the form of the *group theory*. Hausdorff underlines that Poincaré has insisted from the beginning on the structure equations among his symbols and their commutators.[52] He then remarks that his own symbolic proof yields the theorem in the group setting too, but not vice versa—as he stresses again. Nonetheless (in the note at p. 20), Poincaré's 1901 paper is cited as a "beautiful" (literally "schöne") source having some weak relationship with Hausdorff's; later (in the note at p. 40), Poincaré's residue-calculus

[51] It is very hard to find quotations of the papers by Schur, Poincaré, and Pascal (concerning with the product of exponentials) in modern books on Lie groups. To the best of our knowledge, exceptions are: Czyż (1994) (quoting Pascal 1902c and Poincaré 1900) and Duistermaat and Kolk 2000 (quoting Poincaré 1899 and Schur 1891, 1893).

[52] Namely, on identities of the type $[X_i, X_j] = \sum_k c_{ijk} X_k$. In practice, Poincaré is dealing with an actual finite dimensional Lie algebra instead of a set of abstract non-commuting indeterminates, in which Hausdorff was interested.

technique is recalled, and a couple of formulas from Poincaré (1900) is compared with Hausdorff's.
- Hausdorff thinks that Pascal's works on the product of two exponentials have not overcome the inconvenience of an abundance of computations. According to him, these massive computations make it hard and difficult to check the validity of Pascal's proof and the applicability to the group theory. He then faults Pascal for omitting—in his last paper—the convergence question.
- A comment on Schur's contribution is missing in Hausdorff's foreword. Schur is acknowledged only at p. 34, for having proved for the first time a certain series expansion involving Bernoulli numbers.

In Bourbaki's book (1972, Chapter III, Historical Note), we find the following comments:

- Poincaré studied the identity $e^U e^V = e^W$ as a possibly multiple valued function for W. Besides briefly describing his general residue-calculus approach, the actual results obtained by Poincaré are not mentioned, whereas he is faulted (together with Campbell and Baker) for a lack of clarity about the question whether the commutators are symbolic expressions or elements of some fixed Lie algebra.
- Pascal's name is listed -*en passant*- as some of those who returned to the question of $e^U e^V = e^W$, but nothing is said about his results, and his papers are *not* even mentioned in the list of references. Hausdorff instead is acknowledged as the only "perfectly precise" reliable source.
- Schur's results on the product of exponentials are *not* mentioned: he is recalled only for his proof of the fact that C^2 assumptions on transformation groups are (up to isomorphism) sufficient to get analytic regularity.

We would like to express some opinions on the above comments by Hausdorff and Bourbaki.

(i) Curiously, Hausdorff insists on reasserting over again that his own theorem implies Poincaré's and not vice versa, but eventually his words become more prudent[53]: this legitimately leads us to suppose that Hausdorff had foreseen that Poincaré proved more than it seems. Indeed, in Poincaré (1900), he introduced for the very first time the universal enveloping algebra of a Lie group, but the relevance of this invention was not caught for several decades (see Ton-That and Tran 1999 for related topics), and probably it passed unnoticed also to Hausdorff. Likely, Hausdorff did not consider the possibility that Poincaré's formula $e^U e^V = e^W$ was more than a result in Lie groups of transformations. Indeed,

[53] Indeed Hausdorff (1906, p. 20) says that, even disregarding the fact that his theorem implies Poincaré's and not vice versa, he believes that his own contribution offers such a novelty and simplification that it cannot be considered a superfluous supplement with respect to Poincaré's paper. Then he feels the need to return once more on this subject, and he stresses again (p. 40) that Poincaré proved only a group-theoretic theorem which is not a *full* substitute for his own, as it *seems* to follow from a further argument (the words are now more cautious). He further adds a quite long remark as a confirmation of his thesis. Moreover, he says (p. 41) that he does not want to discuss the possibility that, from the validity of Poincaré's theorem on an r-dimensional group for every arbitrary r, one could eventually get the symbolic theorem. Indeed, he points out that for his scopes it was important only to observe that *at least not immediately* (another prudent expression) Poincaré's theorem does not comprehend his general result.

this formula is a formal identity in the universal enveloping algebra $\mathscr{U}(\mathfrak{g})$ of any finite dimensional Lie algebra $\mathfrak{g}$, which determines W as a Lie series in U, V, and (when specialized to $\mathfrak{g}$) the formal series W converges near $0 \in \mathfrak{g}$. Further, as Schmid (1982, p. 184) neatly pointed out, Poincaré's theorem *makes no reference* at all to $\mathfrak{g}$ being the Lie algebra of the infinitesimal transformations of a group.[54] It then seems that Bourbaki's and Hausdorff's criticisms concerning this fact should be mitigated.

(ii) Hausdorff complains about the abundance of computations in Pascal's papers and he says (see Hausdorff 1906, p. 20, lines 21–22) that these massive calculations make it very difficult to check the correctness of Pascal's proof. In our opinion, the presence of abundant computations should not be enough to despise a proof. In this case, the main criticism may concern the fact that Pascal focused on groups of transformations; however, we believe that the greatest part of Pascal's arguments in his series of papers (Pascal 1901a,b, 1902a,b,c) are, though long and laborious, more transparent than some of the proofs only sketched by Campbell or Baker.

(iii) Even though Bourbaki's historical notes do not pay attention to their contributions, Pascal and Schur should be regarded as noteworthy precursors of what was lately to be known only as the Campbell–(Baker–)Hausdorff Formula. Following Schmid (1982, p. 177), this formula is already contained "in disguise" in Schur's papers. Pascal contributed in disclosing this fact (see §2 in Pascal 1902c, p. 565) by proving that the same formulas found by Schur may be obtained starting from his own version of the Exponential Theorem, a completely undisguised result.

Acknowledgments The authors thank Professors Umberto Bottazzini and Jeremy Gray for their careful reading of a previous version of the manuscript, and are especially grateful to the latter for having brought to their attention Scharlau's commentary in Hausdorff (2001), from which they greatly benefited.

References

Abbaspour, H., and M. Moskowitz. 2007. *Basic Lie theory*. Hackensack, NJ: World Scientific.

Alexandrov, P.S., et al. 1979. *Die Hilbertschen Probleme*. Ostwalds Klassiker der exakten Wissenschaften. Leipzig: Akademische Verlagsgesellschaft Geest & Portig K.-G.

Baker, H.F. 1901. On the exponential theorem for a simply transitive continuous group, and the calculation of the finite equations from the constants of structure. *Proceedings of the London Mathematical Society* 34: 91–127.

Baker, H.F. 1902. Further applications of matrix notation to integration problems. *Proceedings of the London Mathematical Society* 34:347–360.

[54] Obviously, it is not possible not to recognize Hausdorff's merits in shifting all the attention to purely symbolic arguments. Nonetheless, our opinion is that Poincaré's ingenious approach (take a finite dimensional Lie algebra $\mathfrak{g}$, obtain the Exponential Theorem on $\mathscr{U}(\mathfrak{g})$, go back to $\mathfrak{g}$ and prove the convergence with the residue calculus) is definitely more modern than a 1900 publication. Furthermore, we think that the only obstacle in obtaining from Poincaré's results the general version of Hausdorff's theorem is the finite-dimensionality of the algebra $\mathfrak{g}$ (as Hausdorff had foreseen!). Indeed, nowadays, we know that, if $\mathfrak{g}$ is the free Lie algebra of a vector space V (note that $\mathfrak{g}$ is infinite-dimensional whenever $V \neq \{0\}$), then $\mathscr{U}(\mathfrak{g})$ is isomorphic to the tensor algebra $\mathscr{T}(V)$: hence, a suitable Exponential Theorem on $\mathscr{U}(\mathfrak{g}) = \mathscr{T}(V)$ may provide the general statement of the CBHD Theorem, which is indeed an identity between formal power series related to $\mathscr{T}(V)$.

Baker, H.F. 1903. On the calculation of the finite equations of a continuous group. *Proceedings of the London Mathematical Society* 35: 332–333.

Baker, H.F. 1905. Alternants and continuous groups. *Proceedings of the London Mathematical Society* (2) 3: 24–47.

Białynicki-Birula, I., Mielnik, B., and J. Plebański. 1969. Explicit solution of the continuous Baker–Campbell–Hausdorff problem and a new expression for the phase operator. *Annals of Physics* 51: 187–200.

Biermann, K.-R. 1988. *Die Mathematik und ihre Dozenten an der Berliner Universität* 1810–1933. *Stationen auf dem Wege eines mathematischen Zentrums von Weltgeltung*, 2nd improved edn. Berlin: Akademie-Verlag.

Birkhoff, G. 1936. Continuous groups and linear spaces. *Recueil Mathématique [Matematicheskii Sbornik] N.S.* 1(43)5: 635–642.

Birkhoff, G. 1938. Analytic groups. *Transactions of the American Mathematical Society* 43: 61–101.

Blanes, S., and F. Casas. 2004. On the convergence and optimization of the Baker-Campbell-Hausdorff formula. *Linear Algebra and its Applications* 378: 135–158.

Blanes, S., F. Casas, J.A. Oteo, and J. Ros. 1998. Magnus and Fer expansions for matrix differential equations: The convergence problem. *Journal of Physics A: Mathematical and General* 31: 259–268.

Blanes, S., F. Casas, J.A. Oteo, and J. Ros. 2009. The Magnus expansion and some of its applications. *Physics Reports* 470: 151–238.

Bonfiglioli, A., and R. Fulci. 2012. *Topics in noncommutative algebra. The theorem of Campbell, Baker, Hausdorff and Dynkin*. Lecture Notes in Mathematics 2034. Berlin: Springer-Verlag.

Borel, A. 2001. *Essays in the history of Lie groups and algebraic groups*. History of Mathematics 21. Providence/Cambridge: American Mathematical Society/London Mathematical Society.

Bose, A. 1989. Dynkin's method of computing the terms of the Baker–Campbell–Hausdorff series. *Journal of Mathematical Physics* 30: 2035–2037.

Boseck, H., G. Czichowski, and K.-P. Rudolph. 1981. Analysis on topological groups—general Lie theory. Teubner-Texte zur Mathematik 37. Leipzig: BSB B. G. Teubner Verlagsgesellschaft.

Bourbaki, N. 1972. *Éléments de Mathématique. Fasc. XXXVII: Groupes et Algèbres de Lie. Chap. II: Algèbres de Lie libres. Chap. III: Groupes de Lie*. Actualités scientifiques et industrielles, 1349. Paris: Hermann.

Campbell, J.E. 1897a. On a law of combination of operators bearing on the theory of continuous transformation groups. *Proceedings of the London Mathematical Society* 28: 381–390.

Campbell, J.E. 1897b. Note on the theory of continuous groups. *Bulletin of the American Mathematical Society* 4: 407–408.

Campbell, J.E. 1898. On a law of combination of operators (second paper). *Proceedings of the London Mathematical Society* 29: 14–32.

Campbell, J.E. 1903. *Introductory treatise on Lie's theory of finite continuous transformation groups*. Oxford: Clarendon Press.

Cartier, P. 1956. Demonstration algébrique de la formule de Hausdorff. *Bulletin de la Société Mathématique de France* 84: 241–249.

Casas, F. 2007. Sufficient conditions for the convergence of the Magnus expansion. *Journal of Physics A: Mathematical and Theoretical* 40: 15001–15017.

Casas, F., and A. Murua. 2009. An efficient algorithm for computing the Baker–Campbell–Hausdorff series and some of its applications. *Journal of Mathematical Physics* 50: 033513-1-033513-23.

Cohen, A. 1931. *An introduction to the Lie theory of one-parameter groups*. New York: G.E. Stechert & Co.

Czyż, J. 1989. On Lie supergroups and superbundles defined via the Baker–Campbell–Hausdorff formula. *Journal of Geometry and Physics* 6: 595–626.

Czyż, J. 1994. *Paradoxes of measures and dimensions originating in Felix Hausdorff's ideas*. Singapore: World Scientific Publishing Co. Inc.

Day, J., W. So, and R.C. Thompson. 1991. Some properties of the Campbell–Baker–Hausdorff series. *Linear and Multilinear Algebra* 29: 207–224.

Dieudonné, J.A. 1974. *Treatise on analysis, Vol. IV*. Pure and Applied Mathematics 10-IV. New York: Academic Press.

Dieudonné, J.A. 1982. *A panorama of pure mathematics. As seen by N. Bourbaki*. Pure and Applied Mathematics 97. New York: Academic Press.

Djoković, D.Ž. 1975. An elementary proof of the Baker–Campbell–Hausdorff–Dynkin formula. *Mathematische Zeitschrift* 143: 209–211.

Dragt, A.J., and J.M. Finn. 1976. Lie series and invariant functions for analytic symplectic maps. *Journal of Mathematical Physics* 17: 2215–2217.
Duistermaat, J.J., and J.A.C. Kolk. 2000. Lie groups. Universitext. Berlin: Springer.
Dynkin, E.B. 1947. Calculation of the coefficients in the Campbell–Hausdorff formula (Russian). *Doklady Akademii Nauk SSSR (N. S.)* 57: 323–326.
Dynkin, E.B. 1949. On the representation by means of commutators of the series $\log(e^x e^y)$ for noncommutative x and y (Russian). *Matematicheskii Sbornik (N.S.)* 25(67)1: 155–162.
Dynkin, E.B. 1950. Normed Lie algebras and analytic groups. *Uspekhi Matematicheskikh Nauk* 5:1(35): 135–186.
Dynkin, E.B. 2000. *Selected papers of E. B. Dynkin with commentary*. Eds. A.A. Yushkevich, G.M. Seitz, and A.L. Onishchik. Providence: American Mathematical Society.
Eichler, M. 1968. A new proof of the Baker–Campbell–Hausdorff formula. *Journal of the Mathematical Society of Japan* 20: 23–25.
Eisenhart, L.P. 1933. *Continuous groups of transformations*. Princeton: University Press. (Reprinted 1961, New York: Dover Publications.)
Eriksen, E. 1968. Properties of higher-order commutator products and the Baker–Hausdorff formula. *Journal of Mathematical Physics* 9: 790–796.
Folland, G.B. 1975. Subelliptic estimates and function spaces on nilpotent Lie groups. *Arkiv for Matematik* 13: 161–207.
Folland, G.B., and E.M. Stein. 1982. *Hardy spaces on homogeneous groups*. Mathematical Notes 28. Princeton/Tokyo: Princeton University Press/University of Tokyo Press.
Friedrichs, K.O. 1953. Mathematical aspects of the quantum theory of fields. V. Fields modified by linear homogeneous forces. *Communications on Pure and Applied Mathematics* 6: 1–72.
Gilmore, R. 1974. Baker–Campbell–Hausdorff formulas. *Journal of Mathematical Physics* 15: 2090–2092.
Glöckner, H. 2002a. Algebras whose groups of units are Lie groups. *Studia Mathematica* 153: 147–177.
Glöckner, H. 2002b. Infinite-dimensional Lie groups without completeness restrictions. In *Geometry and analysis on finite and infinite-dimensional Lie Groups*, vol. 55, ed. A. Strasburger, W. Wojtynski, J. Hilgert, and K.-H. Neeb, 43–59. Warsaw: Banach Center Publications.
Glöckner, H. 2002c. Lie group structures on quotient groups and universal complexifications for infinite-dimensional Lie groups. *Journal of Functional Analysis* 194: 347–409.
Glöckner, H., and K.-H. Neeb. 2003. Banach–Lie quotients, enlargibility, and universal complexifications. *Journal für die reine und angewandte Mathematik (Crelle's Journal)* 560:1–28.
Godement, R. 1982. *Introduction à la théorie des groupes de Lie. Tome 2*. Publications Mathématiques de l'Université Paris VII. Paris: Université de Paris VII, U.E.R. de Mathématiques.
Gorbatsevich, V.V., A.L. Onishchik, and E.B. Vinberg. 1997. *Foundations of Lie theory and Lie transformation groups*. New York: Springer.
Gordina, M. 2005. Hilbert–Schmidt groups as infinite-dimensional Lie groups and their Riemannian geometry. *Journal of Functional Analysis* 227: 245–272.
Hairer, E., Ch. Lubich, and G. Wanner. 2006. *Geometric numerical integration. Structure-preserving algorithms for ordinary differential equations*. New York: Springer.
Hall, B.C. 2003. *Lie groups, Lie algebras, and representations: an elementary introduction*. Graduate texts in mathematics. New York: Springer.
Hausdorff, F. 1906. Die symbolische Exponentialformel in der Gruppentheorie. *Berichte der Königlich-Sächsischen Gesellschaft der Wissenschaften zu Leipzig (Leipziger Berichte), Math. Phys. Cl.* 58: 19–48.
Hausdorff, F. 2001. Gesammelte Werke. Band IV: Analysis, Algebra und Zahlentheorie. Herausgegeben von S.D. Chatterji, R. Remmert und W. Scharlau. (Collected works. Vol. IV: Analysis, algebra, number theory.). Eds. S.D. Chatterji, R. Remmert, and W. Scharlau. Berlin: Springer.
Hausner, M., and J.T. Schwartz. 1968. *Lie Groups. Lie algebras*. Notes on mathematics and its applications. New York: Gordon and Breach.
Hawkins, Th. 1989. Line geometry, differential equations and the birth of Lie's theory of groups. In *The history of modern mathematics*, vol. I, pp. 275–327. Boston: Academic Press.
Hawkins, Th. 1991. Jacobi and the birth of Lie's theory of groups. *Archive for History of Exact Sciences* 42: 187–278.
Hawkins, Th. 2000. *Emergence of the Theory of Lie groups. An essay in the history of mathematics 1869–1926*. Sources and studies in the history of mathematics and physical sciences. New York: Springer.
Hilgert, J., and K.H. Hofmann. 1986. On Sophus Lie's fundamental theorem. *Journal of Functional Analysis* 67: 239–319.

Hilgert, J., and K.-H. Neeb. 1991. *Lie-Gruppen und Lie-Algebren*. Braunschweig: Vieweg.
Hochschild, G.P. 1965. *The structure of Lie groups*. Francisco: San Holden-Day Inc.
Hofmann, K.H. 1972. Die Formel von Campbell, Hausdorff und Dynkin und die Definition Liescher Gruppen. In *Theory of sets and topology*, ed. W. Rinow, 251–264. Berlin: VEB Deutscher Verlag der Wissenschaften.
Hofmann, K.H. 1975. Théorie directe des groupes de Lie. I, II, III, IV. Séminaire Dubreil, Algèbre, tome 27, n. 1 (1973/1974), Exp. 1 (1–24), Exp. 2 (1–16), Exp. 3 (1–39), Exp. 4 (1–15).
Hofmann, K.H., and S.A. Morris. 2005. Sophus Lie's third fundamental theorem and the adjoint functor theorem. *Journal of Group Theory* 8: 115–133.
Hofmann, K.H., and S.A. Morris. 2006. *The structure of compact groups. A primer for the student—A handbook for the expert*, 2nd revised edn., de Gruyter Studies in Mathematics 25. Berlin: Walter de Gruyter.
Hofmann, K.H., and S.A. Morris. 2007. *The Lie theory of connected pro-Lie goups. A structure theory for pro-Lie algebras, pro-Lie groups, and connected locally compact groups*. EMS tracts in mathematics 2. Zürich: European Mathematical Society.
Hofmann, K.H., and K.-H. Neeb. 2009. Pro-Lie groups which are infinite-dimensional Lie groups. *Mathematical Proceedings of the Cambridge Philosophical Society* 146: 351–378.
Hörmander, L. 1967. Hypoelliptic second order differential equations. *Acta Mathematica* 119: 147–171.
Ince, E.L. 1927. *Ordinary differential equations*. London: Longmans, Green & Co. (Reprinted 1956, New York: Dover Publications Inc.)
Iserles, A., H.Z. Munthe-Kaas, S.P. Nørsett, and A. Zanna. 2000. Lie-group methods. *Acta Numerica* 9: 215–365.
Iserles, A., and S.P. Nørsett. 1999. On the solution of linear differential equations in Lie groups. *Philosophical Transactions of the Royal Society A* 357: 983–1019.
Jacobson, N. 1962. *Lie algebras*. Interscience tracts in pure and applied mathematics 10. New York: Interscience Publishers/Wiley.
Klarsfeld, S., and J.A. Oteo. 1989a. Recursive generation of higher-order terms in the Magnus expansion. *Physical Review A* 39: 3270–3273.
Klarsfeld, S., and J.A. Oteo. 1989b. The Baker–Campbell–Hausdorff formula and the convergence of Magnus expansion. *Journal of Physics A: Mathematical and General* 22: 4565–4572.
Kobayashi, H., N. Hatano, and M. Suzuki. 1998. Goldberg's theorem and the Baker–Campbell–Hausdorff formula. *Physica A* 250: 535–548.
Kolsrud, M. 1993. Maximal reductions in the Baker–Hausdorff formula. *Journal of Mathematical Physics* 34: 270–286.
Kumar, K. 1965. On expanding the exponential. *Journal of Mathematical Physics* 6: 1928–1934.
Magnus, W. 1950. A connection between the Baker-Hausdorff formula and a problem of Burnside. *Annals of Mathematics* 52: 111–126.
Magnus, W., A. Karrass, and D. Solitar. 1966. *Combinatorial group theory*. New York: Interscience.
McLachlan, R.I., and R. Quispel. 2002. Splitting methods. *Acta Numerica* 11: 341–434.
Michel, J. 1976. Calculs dans les algèbres de Lie libre: la série de Hausdorff et le problème de Burnside. *Astérisque* 38: 139–148.
Mielnik, B., and J. Plebański. 1970. Combinatorial approach to Baker–Campbell–Hausdorff exponents. *Annales de l'Institut Henri Poincaré* 12: 215–254.
Moan, P.C. 1998. *Efficient approximation of Sturm-Liouville problems using Lie-group methods*. Technical Report 1998/NA11, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England.
Moan, P.C., and J. Niesen. 2008. Convergence of the Magnus series. *Foundations of Computational Mathematics* 8: 291–301.
Moan, P.C., and J.A. Oteo. 2001. Convergence of the exponential Lie series. *Journal of Mathematical Physics* 42: 501–508.
Montgomery, D., and L. Zippin. 1955. *Topological transformation groups*. New York: Interscience Publishers.
Murray, F.J. 1962. Perturbation theory and Lie algebras. *Journal of Mathematical Physics* 3: 451–468.
Nagel, A., E.M. Stein, and S. Wainger. 1985. Balls and metrics defined by vector fields I, basic properties. *Acta Mathematica* 155: 103–147.
Neeb, K.-H. 2006. Towards a Lie theory of locally convex groups. *Japanese Journal of Mathematics* (3) 1: 291–468.

Newman, M., W. So, and R.C. Thompson. 1989. Convergence domains for the Campbell–Baker–Hausdorff formula. *Linear Multilinear Algebra* 24: 301–310.

Omori, H. 1997. *Infinite-dimensional Lie groups*. Translations of Mathematical Monographs 158. Providence: American Mathematical Society.

Oteo, J.A. 1991. The Baker-Campbell-Hausdorff formula and nested commutator identities. *Journal of Mathematical Physics* 32: 419–424.

Pascal, E. 1901a. Sopra alcune indentità fra i simboli operativi rappresentanti trasformazioni infinitesime. *Rendiconti/Istituto Lombardo di Scienze e Lettere, Milano* (2) 34: 1062–1079.

Pascal, E. 1901b. Sulla formola del prodotto di due trasformazioni finite e sulla dimostrazione del cosidetto secondo teorema fondamentale di Lie nella teoria dei gruppi. *Rendiconti/Istituto Lombardo di Scienze e Lettere, Milano* (2) 34: 1118–1130.

Pascal, E. 1902a. Sopra i numeri bernoulliani. *Rendiconti/Istituto Lombardo di Scienze e Lettere, Milano* (2) 35: 377–389.

Pascal, E. 1902b. Del terzo teorema di Lie sull'esistenza dei gruppi di data struttura. *Rendiconti/Istituto Lombardo di Scienze e Lettere, Milano* (2) 35: 419–431.

Pascal, E. 1902c. Altre ricerche sulla formola del prodotto di due trasformazioni finite e sul gruppo parametrico di un dato. *Rendiconti / Istituto Lombardo di Scienze e Lettere, Milano* (2) 35: 555–567.

Pascal, E. 1903. *I Gruppi Continui di Trasformazioni (Parte generale della teoria).* Manuali Hoepli, Nr. 327 bis 328; Milano: Hoepli.

Poincaré, H. 1899. Sur les groupes continus. *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique* 128: 1065–1069.

Poincaré, H. 1900. Sur les groupes continus. *Transactions of the Cambridge Philosophical Society* 18: 220–255.

Poincaré, H. 1901. Quelques remarques sur les groupes continus. *Rendiconti del Circolo Matematico di Palermo* 15: 321–368.

Reinsch, M.W. 2000. A simple expression for the terms in the Baker–Campbell–Hausdorff series. *Journal of Mathematical Physics* 41: 2434–2442.

Reutenauer, C. 1993. *Free Lie algebras*. London Mathematical Society Monographs (New Series) 7. Oxford: Clarendon Press.

Richtmyer, R.D., and S. Greenspan. 1965. Expansion of the Campbell–Baker–Hausdorff formula by computer. *Communications on Pure and Applied Mathematics* 18: 107–108.

Robart, Th. 1997. Sur l'intégrabilité des sous-algèbres de Lie en dimension infinie. *Canadian Journal of Mathematics* 49: 820–839.

Robart, Th. 2004. On Milnor's regularity and the path-functor for the class of infinite dimensional Lie algebras of CBH type. *Algebras, Groups and Geometries* 21: 367–386.

Rossmann, W. 2002. *Lie Groups. An Introduction Through Linear Groups*. Oxford graduate texts in mathematics 5. Oxford: Oxford University Press.

Rothschild, L.P., and E.M. Stein. 1976. Hypoelliptic differential operators and nilpotent groups. *Acta Mathematica* 137: 247–320.

Sagle, A.A., and R.E. Walde. 1973. *Introduction to Lie groups and Lie algebras*. Pure and applied mathematics 51. New York: Academic Press.

Schmid, W. 1982. Poincaré and Lie groups. *Bulletin of the American Mathematical Society (New Series)* 6: 175–186.

Schmid, R. 2010. Infinite-dimensional Lie groups and algebras in Mathematical Physics. *Advances in Mathematical Physics*: Article ID 280362. doi:10.1155/2010/280362.

Schur, F. 1890a. Neue Begründung der Theorie der endlichen Transformationsgruppen. *Mathematische Annalen* 35: 161–197.

Schur, F. 1890b. Beweis für die Darstellbarkeit der infinitesimalen Transformationen aller transitiven endlichen Gruppen durch Quotienten beständig convergenter Potenzreihen. *Berichte der Königlich-Sächsischen Gesellschaft der Wissenschaften zu Leipzig (Leipziger Berichte), Math. phys. Cl.* 42: 1–7.

Schur, F. 1891. Zur Theorie der endlichen Transformationsgruppen. *Mathematische Annalen* 38: 263–286.

Schur, F. 1893. Ueber den analytischen Charakter der eine endliche continuirliche Transformationsgruppe darstellenden Functionen. *Mathematische Annalen* 41: 509–538.

Sepanski, M.R. 2007. *Compact Lie groups*. Graduate texts in mathematics 235. New York: Springer.

Serre, J.P. 1965. *Lie algebras and Lie groups: 1964 lectures given at Harvard University*, 1st edn. 1965, New York: W. A. Benjamin, Inc., 2nd ed. 1992. Lecture notes in mathematics 1500. Berlin: Springer.

Specht, W. 1948. Die linearen Beziehungen zwischen höheren Kommutatoren. *Mathematische Zeitschrift* 51: 367–376.

Suzuki, M. 1977. On the convergence of exponential operators–the Zassenhaus formula, BCH formula and systematic approximants. *Communications in Mathematical Physics* 57: 193–200.

Thompson, R.C. 1982. Cyclic relations and the Goldberg coefficients in the Campbell–Baker–Hausdorff formula. *Proceedings of the American Mathematical Society* 86: 12–14.

Thompson, R.C. 1989. Convergence proof for Goldberg's exponential series. *Linear Algebra and its Applications* 121: 3–7.

Ton-That, T., and T.D. Tran. 1999. Poincaré's proof of the so-called Birkhoff–Witt theorem. *Revue d'Histoire des Mathématiques* 5: 249–284.

Tu, L.W. 2004. Une courte démonstration de la formule de Campbell–Hausdorff. *Journal of Lie Theory* 14: 501–508.

Van Est W.T., and J. Korthagen 1964. Non-enlargible Lie algebras. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen. Series A* 67 = *Indagationes Mathematicae* 26: 15–31.

Varadarajan, V.S. 1984. *Lie groups, Lie algebras, and their representations*. (Reprint of the 1974 edition.) Graduate texts in mathematics 102. New York: Springer.

Vasilescu, F.-H. 1972. Normed Lie algebras. *Canadian Journal of Mathematics* 24: 580–591.

Veldkamp, F.D. 1980. A note on the Campbell–Hausdorff formula. *Journal of Algebra* 62: 477–478.

Wei, J. 1963. Note on the global validity of the Baker–Hausdorff and Magnus theorems. *Journal of Mathematical Physics* 4: 1337–1341.

Weiss, G.H., and T.D. Maradudin. 1962. The Baker Hausdorff formula and a problem in Crystal Physics. *Journal of Mathematical Physics* 3: 771–777.

Wever, F. 1949. Operatoren in Lieschen Ringen. *Journal für die reine und angewandte Mathematik (Crelle's Journal)* 187: 44–55.

Wichmann, E.H. 1961. Note on the algebraic aspect of the integration of a system of ordinary linear differential equations. *Journal of Mathematical Physics* 2: 876–880.

Wilcox, R.M. 1967. Exponential operators and parameter differentiation in quantum physics. *Journal of Mathematical Physics* 8: 962–982.

Wojtyński, W. 1998. Quasinilpotent Banach-Lie algebras are Baker–Campbell–Hausdorff. *Journal of Functional Analysis* 153: 405–413.

Yosida, K. 1937. On the exponential-formula in the metrical complete ring. *Proceedings of the Imperial Academy Tokyo* 13: 301–304. doi:10.3792/pia/1195579861.