TRANSFORMING THE HEUN EQUATION TO THE HYPERGEOMETRIC EQUATION: I. POLYNOMIAL TRANSFORMATIONS

ROBERT S. MAIER*

Abstract. The reductions of the Heun equation to the hypergeometric equation by rational changes of its independent variable are classified. Heun-to-hypergeometric transformations are analogous to the classical hypergeometric identities (i.e., hypergeometric-to-hypergeometric transformations) of Goursat. However, a transformation is possible only if the singular point location parameter and normalized accessory parameter of the Heun equation are each restricted to take values in a discrete set. The possible changes of variable are all polynomial. They include quadratic and cubic transformations, which may be performed only if the singular points of the Heun equation form a harmonic or an equianharmonic quadruple, respectively; and several higher-degree transformations.

Key words. Special functions, Heun equation, hypergeometric equation, hypergeometric identities, Lamé equation, Clarkson–Olver transformation.

AMS subject classifications. 33E30, 34M35, 33C05.

1. Introduction. Consider the class of linear second-order differential equations on the Riemann sphere $\mathbb{CP}^1$ which are Fuchsian, i.e., have only regular singular points [7]. Any such equation with exactly three singular points can be transformed to the hypergeometric equation by appropriate changes of the independent and dependent variables. Similarly, any such equation with exactly four singular points can be transformed to the Heun equation ([6], Chap. 15; [11, 13]).

Solutions of the Heun equation are much less well understood than hypergeometric functions [2]. No general integral representation for them is known, for instance. Most work on solutions of the Heun equation has focused on solutions of special cases, such as the Lamé equation [6, 7]. The parameters of the Heun equation include four characteristic exponent parameters, a singular point location parameter, and a global accessory parameter, so there is a large parameter space to investigate. This is in contrast to the hypergeometric equation, which has only three parameters.

Solutions of Heun equations, such as the Lamé equation, are occasionally used in mathematical modeling. However, it is difficult to carry out practical computations involving them. An explicit solution to the two-point connection problem for the general Heun equation is not known [12], though the corresponding problem for the hypergeometric equation has a classical solution. Determining when the solutions of a Heun equation are expressible in closed form in terms of more familiar functions would obviously be useful: it would facilitate the solution of the two-point connection problem, and the computation of Heun equation monodromies. A significant result in this direction was obtained by Kuiken [9]. It is sometimes possible, by performing a quadratic change of the independent variable, to transform the Heun equation to the hypergeometric equation, and thereby express its solutions in terms of hypergeometric functions. Kuiken's quadratic transformations are not so well known as they should be. The useful monograph edited by Ronveaux [11] does not mention them explicitly, though it lists Ref. [9] in its bibliography. Recently, one of Kuiken's transformations was rediscovered by Ivanov [8].

Unfortunately, the theorem of Ref. [9] is incomplete. The theorem asserts that a transformation to the hypergeometric equation, by a rational change of the independent variable, is possible only if the singular points of the Heun equation form a harmonic quadruple in

*Depts. of Mathematics and Physics, University of Arizona, Tucson AZ 85721. This research was partially supported by NSF grants PHY-9800979 and PHY-0099484.

the sense of projective geometry; in which case the change of variables must be quadratic. We shall show that there are several other possibilities. A transformation may also be possible if the singular points form an equianharmonic quadruple, in which case the change of variables must be cubic. There are additional singular point configurations that permit transformations of degrees 3, 4, 5, and 6. Our main theorem (Theorem 3.1) and its corollaries classify all such transformations, up to linear automorphisms of the Heun and hypergeometric equations.

We were led to this correction and expansion of the theorem of Ref. [9] by considering a discovery of Clarkson and Olver [5]: an unexpected reduction of the Weierstrass form of the equianharmonic Lamé equation to the hypergeometric equation. Their reduction involves a cubic change of the independent variable, which, it turns out, is a special case of a general transformation. In §4, we determine the extent to which their result can be generalized.

Our new transformations are similar to the polynomial transformations which appear in the classical hypergeometric identities (i.e., hypergeometric-to-hypergeometric transformations) of Goursat. (See [6], Chap. 2; also [3].) Transforming the Heun equation to the hypergeometric equation is more difficult: it is possible only if the singular point location parameter and normalized accessory parameter are restricted to take values in related discrete sets. Actually, the Heun-to-hypergeometric transformations classified in this part of the paper (Part I) are of a restricted type: unlike most classical hypergeometric transformations, they include no linear change of the dependent variable. A classification of Heun-to-hypergeometric transformations of the more general type is possible, but is best phrased in geometric terms: it relies on a classification of certain branched covers of the Riemann sphere by itself. Such a classification will appear in Part II.

2. Preliminaries.

2.1. The Equations. The hypergeometric equation may be written in the form

$$\frac{d^2y}{dz^2} + \left(\frac{c}{z} + \frac{a+b-c+1}{z-1}\right)\frac{dy}{dz} + \frac{ab}{z(z-1)}\,y = 0. \tag{$\mathfrak{h}$}$$

It and its solution space are specified by the P-symbol

$$P\left\{\begin{array}{cccc} 0 & 1 & \infty & \\ 0 & 0 & a & ;z \\ 1-c & c-a-b & b & \end{array}\right\}, \tag{2.1}$$

where each column, except the last, refers to a singular point. The first entry is its location, and the final two are the characteristic exponents of the solutions there. The exponents at each singular point are obtained by solving the Frobenius indicial equation [7]. In general, each finite singular point z_0 has ζ as a characteristic exponent if and only if the equation has a solution of the form $(z-z_0)^\zeta h(z)$ in a neighborhood of $z = z_0$, where h is analytic and nonzero at $z = z_0$. If the exponents at $z = z_0$ differ by an integer, this statement must be modified: the solution corresponding to the smaller exponent may have a logarithmic singularity at $z = z_0$. The definition extends in a straightforward way to $z_0 = \infty$, and to ordinary points, each of which is said to have exponents $0, 1$.

There are $2 \times 3 = 6$ local solutions of the hypergeometric equation in all: two per singular point. If c is not a nonpositive integer, the local solution at $z = 0$ belonging to the exponent zero will be analytic. In that case, when normalized to unity at $z = 0$, it will be the Gauss hypergeometric function ${}_2F_1(a, b; c; z)$ [6]. This is the sum of a hypergeometric series, which converges in a neighborhood of $z = 0$. In general, ${}_2F_1(a, b; c; z)$ is not defined when c is a nonpositive integer.

The Heun equation is conventionally written in the form

$$(\mathfrak{H}) \qquad \frac{d^2u}{dt^2} + \left(\frac{\gamma}{t} + \frac{\delta}{t-1} + \frac{\epsilon}{t-d}\right)\frac{du}{dt} + \frac{\alpha\beta t - q}{t(t-1)(t-d)}\,u = 0.$$

Here $d \in \mathbb{C}$, the location of the fourth regular singular point, is a parameter ($d \neq 0, 1$), and the exponent parameters $\alpha, \beta, \gamma, \delta, \epsilon \in \mathbb{C}$ are constrained to satisfy

$$\epsilon = \alpha + \beta - \gamma - \delta + 1. \qquad (2.2)$$

The Heun equation and its solutions are specified by the Riemann P-symbol

$$P\left\{\begin{array}{ccccc} 0 & 1 & d & \infty & \\ 0 & 0 & 0 & \alpha & ;t \\ 1-\gamma & 1-\delta & 1-\epsilon & \beta & \end{array}\right\}. \qquad (2.3)$$

The condition (2.2) constrains the sum of the characteristic exponents to equal 2. The P-symbol does not fully specify the Heun equation and its solutions, since it omits the accessory parameter $q \in \mathbb{C}$.

There are $2 \times 4 = 8$ local solutions of the Heun equation in all: two per singular point. If γ is not a nonpositive integer, the local solution at $t = 0$ belonging to the exponent zero will be analytic. In that case, when normalized to unity at $t = 0$, the solution is called the local Heun function, and is denoted $Hl(d, q; \alpha, \beta, \gamma, \delta; t)$ [11]. It is the sum of a Heun series, which converges in a neighborhood of $t = 0$ [11, 13]. In general, $Hl(d, q; \alpha, \beta, \gamma, \delta; t)$ is not defined when γ is a nonpositive integer.

The solution spaces of the hypergeometric equation and Heun equation are 2-dimensional vector spaces of global analytic functions, i.e., $\mathbb{CP}^1$-valued functions on the multiply punctured Riemann sphere $\mathbb{CP}^1 \setminus \{0, 1, \infty\}$ (resp. $\mathbb{CP}^1 \setminus \{0, 1, d, \infty\}$). In general, the solutions are multivalued. They may be viewed as single-valued analytic functions on a parameter-dependent Riemann surface $\mathcal{S}$ over $\mathbb{CP}^1$, branched at $z = 0, 1, \infty$ (resp. $t = 0, 1, d, \infty$). For certain parameter choices, $\mathcal{S}$ may be compact, in which case the solutions, projected to the multiply punctured Riemann sphere, will be finite-valued, i.e., algebraic. We mention this only in passing: Heun-to-hypergeometric transformations will turn out to exist even in the absence of algebraicity.

If $\epsilon = 0$ and $q = \alpha\beta d$, the Heun equation loses a singular point and becomes the hypergeometric equation. Similar losses occur if $\delta = 0$, $q = \alpha\beta$, or $\gamma = 0$, $q = 0$. Our results will exclude the degenerate case when the Heun equation has fewer than four singular points, since transforming the hypergeometric equation into itself is a separate problem, leading to hypergeometric identities. Also, our treatment will initially exclude the following case, in which $(\mathfrak{H})$ can be solved by elementary means.

DEFINITION 2.1. *If $\alpha\beta = 0$ and $q = 0$, the Heun equation $(\mathfrak{H})$ is said to be trivial. Triviality implies one of the exponents at $t = \infty$ is zero (i.e., $\alpha\beta = 0$), and is implied by absence of the singular point at $t = \infty$ (i.e., $\alpha\beta = 0$, $\alpha + \beta = 1$, $q = 0$).*

The transformation to the Heun equation or hypergeometric equation of a linear second-order Fuchsian differential equation with singular points at $t = 0, 1, d, \infty$ (resp. $z = 0, 1, \infty$), and with arbitrary characteristic exponents, is accomplished by linear changes of the dependent variable, called F-homotopic transformations. (See [6] and [11], §A2 and Addendum, §1.8.) If an equation with singular points at $t = 0, 1, d, \infty$ has dependent variable u, carrying out the substitution $\tilde{u}(t) = t^{-\rho}(t-1)^{-\sigma}(t-d)^{-\tau}u(t)$ will convert the equation to a new one, with the exponents at $t = 0, 1, d$ reduced by ρ, σ, τ respectively, and those at $t = \infty$ increased by $\rho + \sigma + \tau$. By this technique, one of the exponents at each finite singular point can be set to zero, yielding the Heun equation. ρ, σ, τ are not unique: in general, there are two possibilities for each.

In fact, the Heun equation has a $(\mathbb{Z}_2)^3$ group of F-homotopic automorphisms, since at each of $t = 0, 1, d$, the exponents $0, \zeta$ can be shifted to $-\zeta, 0$, which is equivalent to $0, -\zeta$. Similarly, the hypergeometric equation has a $(\mathbb{Z}_2)^2$ group of F-homotopic automorphisms. These groups act on the 6-dimensional and 3-dimensional parameter spaces, respectively. For example, the latter includes $(a, b; c) \mapsto (c - a, c - b; c)$, which is obtained from an F-homotopic transformation at $z = 1$. The identity

$$ {}_2F_1(a,\, b;\, c;\, z) = (1-z)^{c-a-b}\, {}_2F_1(c-a,\, c-b;\, c;\, z) \tag{2.4} $$

is satisfied by ${}_2F_1$, since ${}_2F_1$ is a local solution at $z = 0$, rather than at $z = 1$.

If the regular singular points of the Fuchsian differential equation are arbitrarily placed, transforming it to the hypergeometric or Heun equation will require a Möbius (i.e., projective linear or homographic) transformation, which repositions the singular points to the standard locations. A unique Möbius transformation maps any three distinct points in $\mathbb{CP}^1$ to any other three distinct points; but the same is not true of four points, which is why the Heun equation has d as a free parameter.

2.2. The Cross-Ratio. Our characterization of Heun equations that can be reduced to the hypergeometric equation will employ the cross-ratio orbit of $\{0, 1, d, \infty\}$, defined thus. If A, B, C, D are four distinct points in $\mathbb{CP}^1$, their cross-ratio is

$$ (A, B; C, D) \stackrel{\text{def}}{=} \frac{(C-A)(D-B)}{(D-A)(C-B)} \in \mathbb{CP}^1 \setminus \{0, 1, \infty\}. \tag{2.5} $$

The cross-ratio is is invariant under Möbius transformations; in fact, $(A_1, B_1; C_1, D_1)$ can be mapped to $(A_2, B_2; C_2, D_2)$ iff their cross-ratios are equal. When studying the mapping of unordered point sets, it is necessary to take the action of permutations into account. Permuting A, B, C, D transforms $(A, B; C, D)$ in a well-defined way, yielding an action of the symmetric group S_4 on $\mathbb{CP}^1 \setminus \{0, 1, \infty\}$. $(A, B; C, D)$ is invariant under interchange of the pairs A, B and C, D, and also under simultaneous interchange of the two points in each pair. So each orbit contains no more than $4!/4 = 6$ cross-ratio values. By examination, the six possible actions of S_4 on $s \in \mathbb{CP}^1 \setminus \{0, 1, \infty\}$ are generated by $s \mapsto 1 - s$ and $s \mapsto 1/s$, and the orbit of s comprises

$$ s, \quad 1-s, \quad 1/s, \quad 1/(1-s), \quad s/(s-1), \quad (s-1)/s, \tag{2.6} $$

which may not be distinct. This is called the cross-ratio orbit of s; or, if $s = (A, B; C, D)$, the cross-ratio orbit of the unordered set $\{A, B, C, D\} \subset \mathbb{CP}^1$.

Generic cross-ratio orbits contain six values, but there are two exceptions: an orbit containing exactly three values and an orbit containing exactly two. If $(A, B; C, D)$ equals -1, then the cross-ratio orbit of $\{A, B, C, D\}$ will clearly be $\{-1, 1/2, 2\}$. The value -1 for $(A, B; C, D)$ defines a so-called harmonic configuration: A, B and C, D are said to be harmonic pairs. More generally, if $(A, B; C, D)$ equals any of -1, $1/2$, or 2, i.e., if $\{-1, 1/2, 2\}$ is the cross-ratio orbit of the unordered set $\{A, B, C, D\}$, then $\{A, B, C, D\}$ is said to be a harmonic quadruple. It is easy to see that if C is at infinity and A, B, D are distinct finite points, then A, B and C, D will be harmonic pairs iff D is the midpoint of the line segment $\overline{AB}$. In consequence, $\{A, B, C = \infty, D\}$ will be a harmonic quadruple iff $\{A, B, D\}$ comprises three points which are collinear and equally spaced. So, $\{A, B, C, D\} \subset \mathbb{CP}^1$ will be a harmonic quadruple iff it can be mapped by a Möbius transformation to a set consisting of three equally spaced finite points and the point at infinity; equivalently, to the vertices of a square in $\mathbb{C}$.

The cross-ratio orbit containing exactly two values comprises the two non-real cube roots of -1, i.e., is $\{(1+i\sqrt{3})/2, (1-i\sqrt{3})/2\}$. If this is the cross-ratio orbit of $\{A,B,C,D\}$, $\{A,B,C,D\}$ is said to be an equianharmonic quadruple. If C is at infinity, $\{A,B,C,D\}$ will be an equianharmonic quadruple iff A,B,D are the vertices of an equilateral triangle in $\mathbb{C}$. If $\mathbb{CP}^1$ is interpreted as a sphere via the usual stereographic projection, then by a linear transformation (a special Möbius transformation), this situation reduces to the case when $A,B,C=\infty,D$ are the vertices of a regular tetrahedron. So, $\{A,B,C,D\}\subset\mathbb{CP}^1$ will be an equianharmonic quadruple iff it can be mapped by a Möbius transformation to the vertices of a regular tetrahedron.

Cross-ratio orbits are of two sorts: real orbits such as the harmonic orbit, and non-real orbits such as the equianharmonic orbit. All values in a real orbit are real, and in a non-real orbit, all have a nonzero imaginary part. So, $\{A,B,C,D\}$ will have a specified real orbit as its cross-ratio orbit iff it can be mapped by a Möbius transformation to a set consisting of three specified collinear points in $\mathbb{C}$ and the point at infinity; equivalently, to the vertices of a specified quadrangle (generically irregular) in $\mathbb{C}$. Similarly, it will have a specified non-real orbit as its cross-ratio orbit iff it can be mapped to a set consisting of a specified triangle in $\mathbb{C}$ and the point at infinity; equivalently, to the vertices of a specified tetrahedron (generically irregular) in $\mathbb{CP}^1$.

The cross-ratio orbit of $\{0,1,d,\infty\}$ will be the harmonic orbit iff d equals -1, $1/2$, or 2, and the equianharmonic orbit iff d equals $(1\pm i\sqrt{3})/2$. In contrast, it will be a specified generic orbit iff d takes one of six orbit-specific values.

The cross-ratio orbit of $\{0,1,d,\infty\}$ being a specified orbit is equivalent to its being the same as the cross-ratio orbit of some specified quadruple of the form $\{0,1,D,\infty\}$, i.e., to there being a Möbius transformation that maps $\{0,1,d,\infty\}$ onto $\{0,1,D,\infty\}$. By examination, this occurs iff $\{0,1,d\}$ is mapped onto $\{0,1,D\}$ by some *linear* transformation, i.e., iff the the triangle $\triangle 01d$ is similar to some specified triangle in $\mathbb{C}$. This gives a geometric significance to the allowed values of d.

For the equianharmonic orbit, in which the six values degenerate to two, the triangle may be chosen to be any equilateral triangle. For a real orbit, all values on the cross-ratio orbit are real. But the statement about the orbit being characterized by $\triangle 01d$ being similar to some triangle in $\mathbb{C}$ is still true in a degenerate sense, in which the vertices of the triangles are allowed to be collinear. For example, the cross-ratio orbit of $\{0,1,d,\infty\}$ being the harmonic orbit is equivalent to the 'triangle' $\triangle 01d$ being similar to the 'triangle' $\triangle 012$, i.e., to the unordered set $\{0,1,d\}$ being a set consisting of three equally spaced collinear points. This will hold iff d equals -1, $1/2$, or 2, in agreement with the definition of a harmonic quadruple.

2.3. Automorphisms. According to the theory of the Riemann P-function, a Möbius transformation M of the independent variable will preserve characteristic exponents. For the hypergeometric equation $(\mathfrak{h})$, this implies that if M is one of the $3!$ Möbius transformations that permute the singular points $z=0,1,\infty$ (i.e., elements of the symmetric group S_3), the exponents of the transformed equation $(\tilde{\mathfrak{h}})$ at its singular points $M(0),M(1),M(\infty)$ will be those of $(\mathfrak{h})$ at $0,1,\infty$. But if M is not linear, i.e., $M(\infty)\neq\infty$, then $(\tilde{\mathfrak{h}})$ will not in general be a hypergeometric equation, since its exponents at $M(\infty)$ may both be nonzero. To convert $(\tilde{\mathfrak{h}})$ to a hypergeometric equation, each permutation in S_3 must in general be followed by an F-homotopic transformation of the form $\tilde{y}(z)=[z-M(\infty)]^{-a}y(z)$ or $\tilde{y}(z)=[z-M(\infty)]^{-b}y(z)$.

DEFINITION 2.2. $\mathrm{Aut}(\mathfrak{h})$, *the automorphism group of the hypergeometric equation* $(\mathfrak{h})$, *is the group of changes of variable (Möbius of the independent variable, linear of the depen-*

dent variable) which leave $(\mathfrak{h})$ *invariant, except for changes of parameter. Similarly,* $\mathrm{Aut}(\mathfrak{H})$ *is the automorphism group of the Heun equation* $(\mathfrak{H})$.

$\mathrm{Aut}(\mathfrak{h})$ acts on the 3-dimensional parameter space of $(\mathfrak{h})$. It contains the group S_3 of permutations of singular points as a subgroup, and the group $(\mathbb{Z}_2)^2$ of F-homotopic transformations as a normal subgroup. So $\mathrm{Aut}(\mathfrak{h}) \simeq (\mathbb{Z}_2)^2 \ltimes S_3$, a semidirect product.

DEFINITION 2.3. *Within* $\mathrm{Aut}(\mathfrak{h})$, *the Möbius automorphism group is the subgroup* $\mathcal{M}(\mathfrak{h}) \stackrel{\text{def}}{=} \{1\} \times S_3$, *which permutes the singular points* $z = 0, 1, \infty$. *The linear automorphism group is* $\mathcal{L}(\mathfrak{h}) \stackrel{\text{def}}{=} \{1\} \times S_2$, *which permutes the* finite *singular points* $z = 0, 1$, *and fixes* $z = \infty$. *The F-homotopic automorphism group is* $(\mathbb{Z}_2)^2 \times \{1\}$.

The action of $\mathrm{Aut}(\mathfrak{h})$ on the $2 \times 3 = 6$ local solutions is as follows. $|\mathrm{Aut}(\mathfrak{h})| = 2^2 \times 3! = 24$, and applying the transformations in $\mathrm{Aut}(\mathfrak{h})$ to any single local solution yields 24 solutions of $(\mathfrak{h})$. Applying them to ${}_2F_1$, for instance, yields the 24 well-known series solutions of Kummer. However, the 24 solutions split into six sets of four, since for each singular point $z_0 \in \{0, 1, \infty\}$ there is a subgroup of $\mathrm{Aut}(\mathfrak{h})$ of order $2^1 \times 2! = 4$, each element of which fixes $z = z_0$ and performs no F-homotopy there; so it leaves each local solution at $z = z_0$ invariant.

For example, the four transformations in the subgroup associated to $z = 0$ yield four equivalent expressions for ${}_2F_1(a, b; c; z)$; one of which is ${}_2F_1(a, b; c; z)$ itself, and another of which appears above in (2.4). The remaining expressions for ${}_2F_1(a, b; c; z)$ are $(1 - z)^{-a}{}_2F_1\,(a, c - b; c; z/(z - 1))$ and $(1 - z)^{-b}{}_2F_1\,(b, c - a; c; z/(z - 1))$. The five additional sets of four are expressions for the five additional local solutions. One which will play a role is the local solution at $z = 0$ belonging to the exponent $1 - c$. One of the four expressions for it, in terms of ${}_2F_1$, is [6]

$$\widetilde{{}_2F_1}(a,\, b;\, c;\, z) \stackrel{\text{def}}{=} z^{1-c}{}_2F_1(a - c + 1,\, b - c + 1;\, 2 - c;\, z). \tag{2.7}$$

The quantity $\widetilde{{}_2F_1}(a, b; c; z)$ is defined if $c \neq 2, 3, 4, \ldots$ The second local solution at $z = 0$ must be specified differently if $c = 2, 3, 4, \ldots$ (See [1], §15.5.) The same is true if $c = 1$, since in that case, $\widetilde{{}_2F_1}$ reduces to ${}_2F_1$. When $\widetilde{{}_2F_1}$ is defined, we define it uniquely in a neighborhood of $z = 0$ by choosing the principal branch of z^{1-c}.

The automorphism group of the Heun equation is slightly more complicated to describe. There are 4! Möbius transformations M that map the singular points $t = 0, 1, d, \infty$ onto $t = 0, 1, d', \infty$, for some $d' \in \mathbb{CP}^1 \setminus \{0, 1, \infty\}$. The possible d' constitute the cross-ratio orbit of $\{0, 1, d, \infty\}$. Of these 4! transformations, 3! fix $t = \infty$, i.e., are linear. All values d' on the orbit are obtained via linear transformations, i.e., a mapping is possible iff $\triangle 01d$ is similar to $\triangle 01d'$. If M is not linear, it must be followed by an F-homotopic transformation of the form $\tilde{u}(t) = [t - M(\infty)]^{-\alpha}u(t)$ or $\tilde{u}(t) = [t - M(\infty)]^{-\beta}u(t)$.

$\mathrm{Aut}(\mathfrak{H})$ acts on the 6-dimensional parameter space of $(\mathfrak{H})$. It contains the group S_4 of permutations of singular points as a subgroup, and the group $(\mathbb{Z}_2)^3$ of F-homotopic transformations as a normal subgroup. So $\mathrm{Aut}(\mathfrak{H}) \simeq (\mathbb{Z}_2)^3 \ltimes S_4$, a semidirect product.

DEFINITION 2.4. *Within* $\mathrm{Aut}(\mathfrak{H})$, *the Möbius automorphism group is the subgroup* $\mathcal{M}(\mathfrak{H}) \stackrel{\text{def}}{=} \{1\} \times S_4$, *which maps between sets of singular points of the form* $\{0, 1, d', \infty\}$. *The linear automorphism group is* $\mathcal{L}(\mathfrak{H}) \stackrel{\text{def}}{=} \{1\} \times S_3$, *which maps between sets of* finite *singular points of the form* $\{0, 1, d'\}$, *and fixes* $t = \infty$. *The F-homotopic automorphism group is* $(\mathbb{Z}_2)^3 \times \{1\}$.

The action of $\mathrm{Aut}(\mathfrak{H})$ on the $2 \times 4 = 8$ local solutions is as follows. $|\mathrm{Aut}(\mathfrak{H})| = 2^3 \times 4! = 192$, and applying the transformations in $\mathrm{Aut}(\mathfrak{H})$ to any single local solution yields 192 solutions of $(\mathfrak{H})$. However, the 192 solutions split into eight sets of 24, since for each singular point $t_0 \in \{0, 1, d, \infty\}$ there is a subgroup of $\mathrm{Aut}(\mathfrak{H})$ of order $2^2 \times 3! = 24$, each element of which fixes $t = t_0$ and performs no F-homotopy there; so it leaves each local solution at $t = t_0$ invariant. This statement must be interpreted with care: in the event that $t_0 = d$, what is really being chosen is a cross-ratio orbit, rather than a single point.

The case $t_0 = 0$ should serve as an example. The 24 transformations in the subgroup associated to $t = 0$ yield 24 equivalent expressions for $Hl(d, q; \alpha, \beta, \gamma, \delta; t)$, one of which, the only nontrivial one with no F-homotopic factor, is [11, 13]

$$Hl\,(d,\,q;\,\alpha,\,\beta,\,\gamma,\,\delta;\,t) = Hl\,(1/d,\,q/d;\,\alpha,\,\beta,\,\gamma,\,\alpha+\beta-\gamma-\delta+1;\,t/d)\,. \tag{2.8}$$

(The two sides are defined if γ is not a nonpositive integer.) The additional seven sets of 24 are expressions for the additional seven local solutions. One which will play a role is the solution at $t = 0$ belonging to the exponent $1-\gamma$. One of the 24 expressions for it, in terms of Hl, is [13]

$$\widetilde{Hl}(d,\,q;\,\alpha,\,\beta,\,\gamma,\,\delta;\,t) \stackrel{\text{def}}{=} t^{1-\gamma}\,Hl(d,\,q';\,\alpha-\gamma+1,\,\beta-\gamma+1,\,2-\gamma,\,\delta;\,t), \tag{2.9}$$

where the transformed accessory parameter q' equals $q + (1-\gamma)(\epsilon + d\delta)$. The quantity $\widetilde{Hl}(d, q; \alpha, \beta, \gamma, \delta; t)$ is defined if $\gamma \neq 2, 3, 4, \dots$ The second local solution at $t = 0$ must be specified differently if $\gamma = 2, 3, 4, \dots$ The same is true if $\gamma = 1$, since in that case, $\widetilde{Hl}$ reduces to Hl. When $\widetilde{Hl}$ is defined, we define it uniquely in a neighborhood of $t = 0$ by choosing the principal branch of $t^{1-\gamma}$.

In general, automorphisms of $(\mathfrak{H})$ will alter not merely d and the exponent parameters, but also the accessory parameter q. This is illustrated by (2.8) and (2.9).

3. Polynomial Heun-to-Hypergeometric Transformations. We can now state, and prove, our corrected and expanded version of the theorem of Ref. [9].

The theorem will characterize when a homomorphism of rational substitution type from the Heun equation $(\mathfrak{H})$ to the hypergeometric equation $(\mathfrak{h})$ exists. It will list the possible substitutions, up to linear automorphisms of the two equations. It is really a statement about which $\mathcal{L}(\mathfrak{H})$-orbits may be mapped by homomorphisms of this type to $\mathcal{L}(\mathfrak{h})$-orbits. The possible substitutions, it turns out, are all polynomial.

For ease of understanding, the characterization of the theorem will be concrete: it will require that the triangle $\triangle 01d$ be similar to one of a set of triangles of the form $\triangle 01D$. Similarity occurs iff d belongs to the cross-ratio orbit of D, i.e., iff d can be generated from D by repeated application of $d \mapsto 1-d$ and $d \mapsto 1/d$. It is worth noting that if $\operatorname{Re} D = 1/2$, the orbit of D is closed under complex conjugation.

For each value of D, the polynomial mapping from $t \in \mathbb{CP}^1$ to $z \in \mathbb{CP}^1$, which we denote R, will be given explicitly when $d = D$. When $d \neq D$, the mapping can be computed by composing with the unique linear transformation L_1 of $\mathbb{C}$ that maps $\triangle 01d$ to $\triangle 01D$. When $d \neq D$, statements about necessary conditions regarding the singular points, characteristic exponents, and accessory parameter must also be modified. For example, case 1 of the theorem refers to a distinguished singular point d_0, the mandatory value of which is given when $d = D$. When $d \neq D$, its mandatory value can be computed as the preimage of that point under L_1. Similarly, a statement like “the characteristic exponents of $t = 0$ are required to be $0, 1/2$”, valid when $d = D$, would be interpreted when $d \neq D$ as “the characteristic exponents of the preimage of $t = 0$ under L_1 are required to be $0, 1/2$”.

Note that in the statement of the theorem and what follows, $S \stackrel{\text{def}}{=} 1 - R$.

THEOREM 3.1. *A Heun equation of the form* $(\mathfrak{H})$*, which has four singular points and is nontrivial (i.e.,* $\alpha\beta \neq 0$ *or* $q \neq 0$*), can be transformed to a hypergeometric equation of the form* $(\mathfrak{h})$ *by a rational substitution* $z = R(t)$ *if and only if* R *is a polynomial,* $\alpha\beta \neq 0$*, and one of the following two conditions is satisfied.*

1. *$\triangle 01d$ is similar to $\triangle 01D$, for one of the values of D listed in subcases* 1(a)–1(c)*; each of which is real, so the triangle must be degenerate. Also, the normalized accessory parameter $q/\alpha\beta$ must equal one of $0, 1, d$, which may be denoted d_0. The description of each subcase lists the value of d_0 when $d = D$.*

2. *$\triangle 01d$ is similar to $\triangle 01D$, for one of the values of D listed in subcases* 2(a)–2(d)*; each of which is non-real and has real part equal to $1/2$, so the triangle must be isosceles. The description of each subcase lists the value of $q/\alpha\beta$ when $d = D$.*

Besides specifying D and the value of $q/\alpha\beta$ when $d = D$, each subcase imposes restrictions on the characteristic exponent parameters at the singular points $0, 1, d$. The subcases of case 1 are the following.

1(a). *[Harmonic (equally spaced collinear points) case.] $D = 2$. Suppose $d = D$. Then d_0 must equal 1, and $t = 0, d$ must have the same characteristic exponents, i.e., $\gamma = \epsilon$. In general, either R or S will be the degree-2 polynomial $t(2-t)$, which maps $t = 0, d$ to $z = 0$ and $t = 1$ to $z = 1$ (with double multiplicity). There are special circumstances in which R may be quartic, which are listed separately, as subcase* 1(c).

1(b). *$D = 4$. Suppose $d = D$. Then d_0 must equal 1, the point $t = 1$ must have characteristic exponents that are double those of $t = d$, i.e., $1 - \delta = 2(1 - \epsilon)$, and $t = 0$ must have exponents $0, 1/2$, i.e., $\gamma = 1/2$. Either R or S will be the degree-3 polynomial $(t-1)^2(1 - t/4)$, which maps $t = 0$ to $z = 1$ and $t = 1, d$ to $z = 0$ (the former with double multiplicity).*

1(c). *[Special harmonic case.] $D = 2$. Suppose $d = D$. Then d_0 must equal 1, and $t = 0, d$ must have the same characteristic exponents, i.e., $\gamma = \epsilon$. Moreover, the exponents of $t = 1$ must be twice those of $t = 0, d$, i.e., $1 - \delta = 2(1-\gamma) = 2(1-\epsilon)$. Either R or S will be the degree-4 polynomial $4[t(2-t) - \frac12]^2$, which maps $t = 0, 1, d$ to $z = 1$ ($t = 1$ with double multiplicity).*

The subcases of case 2 are the following.

2(a). *[Equianharmonic (equilateral triangle) case.] $D = \frac12 + i\sqrt3/2$. $q/\alpha\beta$ must equal the mean of $0, 1, d$, and $t = 0, 1, d$ must have the same characteristic exponents, i.e., $\gamma = \delta = \epsilon$. Suppose $d = D$. Then $q/\alpha\beta$ must equal $\frac12 + i\sqrt3/6$. In general, either R or S will be the degree-3 polynomial $\left[1 - t/(\frac12 + i\sqrt3/6)\right]^3$, which maps $t = 0, 1, d$ to $z = 1$ and $t = q/\alpha\beta$ to $z = 0$ (with triple multiplicity). There are special circumstances in which R may be sextic, which are listed separately, as subcase* 2(d).

2(b). *$D = \frac12 + i5\sqrt2/4$. Suppose $d = D$. Then $q/\alpha\beta$ must equal $\frac12 + i\sqrt2/4$, $t = d$ must have characteristic exponents $0, 1/3$, i.e., $\epsilon = 2/3$, and $t = 0, 1$ must have exponents $0, 1/2$, i.e., $\gamma = \delta = 1/2$. Either R or S will be the degree-4 polynomial $\left[1 - t/(\frac12 + i5\sqrt2/4)\right]\left[1 - t/(\frac12 + i\sqrt2/4)\right]^3$, which maps $t = d, q/\alpha\beta$ to $z = 0$ (the latter with triple multiplicity) and $t = 0, 1$ to $z = 1$.*

2(c). *$D = \frac12 + i11\sqrt{15}/90$. Suppose $d = D$. Then $q/\alpha\beta$ must equal $\frac12 + i\sqrt{15}/18$, $t = d$ must have characteristic exponents $0, 1/2$, i.e., $\epsilon = 1/2$, and $t = 0, 1$ must have exponents $0, 1/3$, i.e., $\gamma = \delta = 2/3$. Either R or S will be the degree-5 polynomial $At(t-1)\left[t - (\frac12 + i\sqrt{15}/18)\right]^3$, which maps $t = 0, 1, q/\alpha\beta$ to $z = 0$ (the last with triple multiplicity). The factor A is chosen so that it maps $t = d$ to $z = 1$, as well; explicitly, $A = -i2025\sqrt{15}/64$.*

2(d). *[Special equianharmonic case.] $D = \frac12 + i\sqrt3/2$. $q/\alpha\beta$ must equal the mean of $0, 1, d$, and $t = 0, 1, d$ must have characteristic exponents $0, 1/3$, i.e., $\gamma = \delta = \epsilon = 2/3$. Suppose $d = D$. Then $q/\alpha\beta$ must equal $\frac12 + i\sqrt3/6$. Either R or S will be the degree-6 polynomial $4\left\{[1 - t/(\frac12 + i\sqrt3/6)]^3 - \frac12\right\}^2$, which maps $t = 0, 1, d, q/\alpha\beta$ to $z = 1$ (the last with triple multiplicity).*

Remark 1. The origin of the special harmonic and equianharmonic subcases is easy to understand. In subcase 1(c), $t \mapsto R(t)$ or $S(t)$ is the composition of the quadratic map of 1(a) with the map $z \mapsto 4(z-\frac{1}{2})^2$. In subcase 2(d), $t \mapsto R(t)$ or $S(t)$ is similarly the composition of the cubic map of 2(a) with $z \mapsto 4(z-\frac{1}{2})^2$. The extra restrictions on exponents make feasible the additional quadratic transformation of z, which is a classical transformation of the hypergeometric equation into itself [6].

Remark 2. R is determined uniquely by the choices enumerated in the theorem. There is a choice of subcase, a choice of d from the cross-ratio orbit of D, and a binary choice between R and S. The final two choices amount to choosing maps $L_1 \in \mathcal{L}(\mathfrak{H})$ and $L_2 \in \mathcal{L}(\mathfrak{h})$, i.e., $L_2(z) = z$ or $1-z$, which precede and follow a canonical substitution.

In the harmonic case 1(a), in which the $\mathcal{L}(\mathfrak{H})$-orbit includes three values of d, there are accordingly $3 \times 2 = 6$ possibilities for R; namely,

$$R = t^2,\ 1-t^2; \quad (2t-1)^2,\ 1-(2t-1)^2; \quad t(2-t),\ 1-t(2-t), \tag{3.1}$$

corresponding to $d = -1, -1; 1/2, 1/2; 2, 2$, respectively. These are the quadratic transformations of Kuiken [9]. In the equianharmonic case 2(a), in which the orbit includes only two values of d, there are $2 \times 2 = 4$ possibilities; namely,

$$R = [1 - t/(\tfrac{1}{2} \pm i\sqrt{3}/6)]^3, \qquad 1 - [1 - t/(\tfrac{1}{2} \pm i\sqrt{3}/6)]^3, \tag{3.2}$$

corresponding to $d = \frac{1}{2} \pm i\sqrt{3}/2$. The remaining subcases, with the exception of 1(c) and 2(d), correspond to generic cross-ratio orbits: each value of D specifies six values of d. In each of those subcases, there are $6 \times 2 = 12$ possibilities. So in all, there are 56 possibilities for R.

Remark 3. The characteristic exponents of the singular points $z = 0, 1, \infty$ follow from those of the singular points $t = 0, 1, d, \infty$, together with the formula for R. The computation relies on Proposition 3.3 below, which may be summarized thus. If $t = t_0$ is not a critical point of the map $t \mapsto z = R(t)$, then the exponents of $z = R(t_0)$ are the same as those of t_0. If, on the other hand, $t = t_0$ is mapped to $z = z_0 \stackrel{\text{def}}{=} R(t_0)$ with multiplicity $k > 1$, i.e., $t = t_0$ is a $k-1$-fold critical point of R and $z = z_0$ is a critical value, then the exponents of z_0 are $1/k$ times those of t_0.

For example, in the harmonic case 1(a), the map $t \mapsto z$ coalesces two of $t = 0, 1, d$ to either $z = 0$ or $z = 1$, and by examination, the coalesced point is not a critical value of the map; so the characteristic exponents of those two points are preserved, and must therefore be the same, as stated in the theorem. On the other hand, the characteristic exponents of the third point of the three, $t = d_0$, are necessarily halved when it is mapped to $z = 1$ or $z = 0$, since by examination, R always has a simple critical point at $t = d_0$, i.e., $z \sim \mathrm{const} + C(t-d_0)^2$ for some nonzero C. (These statements follow by considering the canonical $d = D$ case.) So if δ_0 denotes the parameter (out of γ, δ, ϵ) corresponding to $t = d_0$, the characteristic exponents of $z = 1$ or $z = 0$ will be $0, (1-\delta_0)/2$. R, being a quadratic polynomial, also has a simple critical point at $t = \infty$, so the characteristic exponents of $z = \infty$ are one-half those of $t = \infty$, i.e., $\alpha/2, \beta/2$. It follows that in the harmonic case, the Gauss parameters $(a, b; c)$ of the resulting hypergeometric equation will be $(\alpha/2, \beta/2; (\delta_0 + 1)/2)$ or $(\alpha/2, \beta/2; (\alpha + \beta - \delta_0 + 1)/2)$.

In the equianharmonic case 2(a), the map $t \mapsto z$ coalesces $t = 0, 1, d$ to either $z = 0$ or $z = 1$; and by examination, the coalesced point is not a critical value of the map; so the characteristic exponents of those three points are preserved, and must therefore be the same, as stated in the theorem. On the other hand, at $t = q/\alpha\beta$, which is mapped to $z = 1$ or $z = 0$, R has, by examination, a double critical point, i.e., $z \sim \mathrm{const} + C(t - q/\alpha\beta)^3$

for some nonzero C. So the characteristic exponents of $z=1$ or $z=0$, since $t=q/\alpha\beta$ is an ordinary point of the Heun equation and effectively has characteristic exponents $0,1$, are $0,1/3$. R, being a cubic polynomial, also has a double critical point at $t=\infty$, so the characteristic exponents of $z=\infty$ are one-third those of $t=\infty$, i.e., $\alpha/3,\beta/3$. It follows that in the equianharmonic case, the parameters $(a,b;c)$ of the resulting hypergeometric equation will be $(\alpha/3,\beta/3;2/3)$ or $(\alpha/3,\beta/3;(\alpha+\beta+1)/3)$.

DEFINITION 3.2. *A rational map $R:\mathbb{CP}^1\to\mathbb{CP}^1$ is said to map the characteristic exponents of the Heun equation $(\mathfrak{H})$ to the characteristic exponents of the hypergeometric equation $(\mathfrak{h})$ if, for all $t_0\in\mathbb{CP}^1$, the exponents of $t=t_0$ according to the Heun equation, divided by the multiplicity of $t_0\mapsto z_0\stackrel{\text{def}}{=}R(t_0)$, equal the exponents of $z=z_0$ according to the hypergeometric equation.*

For example, if t_0 and z_0 are both finite, this says that if $z\sim z_0+C(t-t_0)^k$ to leading order, for some nonzero C, then the exponents of $z=z_0$ must be those of $t=t_0$, divided by k. If $t=t_0$ is an ordinary point of the Heun equation, then the exponents of $z=z_0$ will be $0,1/k$. This implies that if $k>1$, $z=z_0$ must be a singular point of the hypergeometric equation, rather than an ordinary point.

PROPOSITION 3.3. *A Heun equation of the form $(\mathfrak{H})$ will be transformed to a hypergeometric equation of the form $(\mathfrak{h})$ by a specified rational substitution $z=R(t)$ of its independent variable only if R maps exponents to exponents.*

Proposition 3.3, which was already used in Remark 3 above, is an immediate consequence of the following lemma, which has an elementary proof. But the proposition is best seen as a special case of a basic fact in the theory of the Riemann P-function: if a rational change of the independent variable transforms one Fuchsian equation on the Riemann sphere to another, then the characteristic exponents are transformed multiplicatively. In that context, it would be proved by examining the effects of the change of independent variable on each local (Frobenius) solution.

The lemma begins the study of *sufficient* conditions for the existence of a Heun-to-hypergeometric transformation. Finding them requires care, since an accessory parameter is involved. It is useful to perform the substitution $z=R(t)$ explicitly. Substituting $z=R(t)$ into the hypergeometric equation $(\mathfrak{h})$ transforms it to

$$\frac{d^2y}{dt^2}+\left\{-\frac{\ddot R}{\dot R}+\frac{\dot R}{R(1-R)}[c-(a+b+1)R]\right\}\frac{dy}{dt}-\frac{ab\dot R^2}{R(1-R)}\,y=0. \tag{3.3}$$

LEMMA 3.4. *The coefficient of the dy/dt term in the transformed hypergeometric equation (3.3), which may be denoted $W(t)$, equals the coefficient of the du/dt term in the Heun equation $(\mathfrak{H})$, i.e., $\gamma/t+\delta/(t-1)+\epsilon/(t-d)$, if and only if R maps exponents to exponents. That is, the transformation at least partly 'works' if and only if R maps exponents to exponents.*

Proof. Suppose R maps $t=t_0$ to $z=z_0\stackrel{\text{def}}{=}R(t_0)$ with multiplicity k, i.e., to leading order $R(t)\sim z_0+C(t-t_0)^k$; if t_0 and z_0 are finite, that is. By direct computation, the leading behavior of W at $t=t_0$ is the following. In the case when t_0 is finite, $W(t)\sim(1-k)(t-t_0)^{-1}$ if $z_0\neq 0,1,\infty$; $[1-k(1-c)](t-t_0)^{-1}$ if $z_0=0$; $[1-k(c-a-b)](t-t_0)^{-1}$ if $z_0=1$, and $[1-k(a+b)](t-t_0)^{-1}$ if $z_0=\infty$. In the case when $t_0=\infty$, $W(t)\sim(1+k)t^{-1}$ if $z_0\neq 0,1,\infty$; $[1+k(1-c)]t^{-1}$ if $z_0=0$; $[1+k(c-a-b)]t^{-1}$ if $z_0=1$, and $[1+k(a+b)]t^{-1}$ if $z_0=\infty$.

This may be restated as follows. At $t=t_0$, for finite t_0, the leading behavior of W is $W(t)\sim(1-k\eta)(t-t_0)^{-1}$, where k is the multiplicity of $t_0\mapsto z_0\stackrel{\text{def}}{=}R(t_0)$ and η is the sum of the two characteristic exponents of the hypergeometric equation at $z=z_0$; if the coefficient $1 - k\eta$ equals zero then W has no pole at $t = t_0$. Likewise, the leading behavior of W at $t = \infty$ is $W(t) \sim (1 + k\eta)t^{-1}$, where k is the multiplicity of $\infty \mapsto z_0 \stackrel{\mathrm{def}}{=} R(\infty)$ and η is the sum of the two exponents at $z = z_0$; if the coefficient $1 + k\eta$ equals zero then W has a higher-order zero at $t = \infty$.

By the definition of 'mapping exponents to exponents', it follows that the leading behavior of W at $t = t_0$, for all t_0 finite, is of the form $W(t) \sim (1 - \eta')(t - t_0)^{-1}$, and also at $t_0 = \infty$, is of the form $W(t) \sim (1 + \eta')t^{-1}$, where in both cases η' is the sum of the exponents of the Heun equation at $t = t_0$, iff R maps exponents to exponents.

That is, the rational function W has leading behavior γt^{-1} at $t = 0$, $\delta(t-1)^{-1}$ at $t = 1$, $\epsilon(t - d)^{-1}$ at $t = d$, $(1 + \alpha + \beta)t^{-1} = (\gamma + \delta + \epsilon)t^{-1}$ at $t = \infty$, and is regular at all t other than $0, 1, d, \infty$, iff R maps exponents to exponents. ∎

The following propositions characterize when the transformed hypergeometric equation (3.3) is, in fact, the Heun equation ($\mathfrak{H}$).

PROPOSITION 3.5. *A Heun equation of the form* ($\mathfrak{H}$)*, which is trivial (i.e.,* $\alpha\beta = 0$ *and* $q = 0$*) will be transformed to a hypergeometric equation of the form* ($\mathfrak{h}$) *by a specified rational substitution* $z = R(t)$ *of its independent variable if and only if* R *maps exponents to exponents.*

Proof. The 'if' half is new, and requires proof. By Lemma 3.4, the coefficients of dy/dt and du/dt agree iff R maps exponents to exponents, so it suffices to determine whether the coefficients of y and u agree. But by triviality, the coefficient of u in ($\mathfrak{H}$) is zero. Also, $t = \infty$ has zero as one of its exponents, so all points $t \in \mathbb{CP}^1$ have zero as an exponent. By the mapping of exponents to exponents, $z = \infty$ must also have zero as an exponent, i.e., $ab = 0$. So the coefficient of y in (3.3) is also zero. ∎

PROPOSITION 3.6. *A Heun equation of the form* ($\mathfrak{H}$)*, which has four singular points and is nontrivial (i.e.,* $\alpha\beta \neq 0$ *or* $q \neq 0$*), will be transformed to a hypergeometric equation of the form* ($\mathfrak{h}$) *by a specified rational substitution* $z = R(t)$ *of its independent variable if and only if* R *maps exponents to exponents, and moreover,* R *is a polynomial,* $\alpha\beta \neq 0$*, and one of the following two conditions on the normalized accessory parameter* $p \equiv q/\alpha\beta$ *is satisfied.*

1. p *equals one of* $0, 1, d$*. Call this point* d_0*, and the other two singular points* d_1 *and* d_2*. In this case,* d_0 *must be a double zero of* R *or* S*, and each of* d_1, d_2 *must be a simple zero of* R *or* S*.*
2. p *does not equal any of* $0, 1, d$*. In this case, each of* $0, 1, d$ *must be a simple zero of* R *or* S*, and* p *must be a triple zero of either* R *or* S*.*

In both cases, R *and* S *can have no additional simple zeroes or zeroes of order greater than two. Also, if* $1 - c$ *(the nonzero exponent at* $z = 0$*) does not equal* $1/2$*, then* R *can have no additional double zeroes; and if* $c - a - b$ *(the nonzero exponent at* $z = 1$*) does not equal* $1/2$*, then* S *can have no additional double zeroes.*

In both cases, no additional double zero, if any, can be mapped by R *to the point (out of* $z = 0, 1$*) to which* p *is mapped. So additional double zeroes, if any, must all be zeroes of* R*, or all be zeroes of* S*.*

Proof. Like Proposition 3.5, this follows by comparing the transformed hypergeometric equation (3.3) to the Heun equation ($\mathfrak{H}$). By Lemma 3.4, the coefficients of dy/dt and du/dt agree iff R maps exponents to exponents, so it suffices to characterize when the coefficients of y and u agree.

The coefficient of y in (3.3) is to equal the coefficient of u in ($\mathfrak{H}$). It follows that $ab = 0$ is possible iff $\alpha\beta = 0$ and $q = 0$, which is ruled out by nontriviality. So $ab \neq 0$, and equality of the coefficients can hold iff

$$U \equiv \frac{\dot{R}}{R}\,\frac{\dot{S}}{S} = \frac{-(\alpha\beta t - q)/ab}{t(t-1)(t-d)} \equiv \frac{C_0}{t} + \frac{C_1}{t-1} + \frac{C_d}{t-d}, \tag{3.4}$$

where $S \equiv 1 - R$, and C_0, C_1, C_d are certain complex numbers, at least two of which are nonzero.

Both $\dot{R}/R$ and $\dot{S}/S$ are sums of terms of the form $n(t-\lambda)^{-1}$, where n is a nonzero integer and λ is a zero or a pole of R or S. Poles are impossible, since λ is a pole of R iff λ is a pole of S, and there are no double poles on the right-hand side of (3.4). So R must be a polynomial.

By examining the definition of U in terms of R and S, one sees the following is true of any $\lambda \in \mathbb{C}$: if R or S has a simple zero at $t = \lambda$, then U will have a simple pole at $t = \lambda$; if R or S has a double zero at $t = \lambda$, then U will have an ordinary point (non-zero, non-pole) at $t = \lambda$, and if R or S has a zero of order $k > 2$ at $t = \lambda$, then U will have a zero of order $k - 2$ at $t = \lambda$.

Most of what follows is devoted to proving the 'only if' half of the proposition in the light of these facts, by examining the consequences of the equality (3.4). In the final paragraph, the 'if' half will be proved.

There are exactly three ways in which the equality (3.4) can hold.

0. $\alpha\beta = 0$, but due to nontriviality, $q \neq 0$. U has three simple poles on $\mathbb{C}$, at $t = 0, 1, d$. It has no other poles, and no zeroes. So each of $0, 1, d$ must be a simple zero of either R or S; also, R and S can have no other simple zeroes, and no zeroes of order $k > 2$. Except for possible double zeroes, the zeroes of R and S are determined. The degree of R must equal the number of zeroes of R, and also equal the number of zeroes of S, counting multiplicity. But irrespective of how many double zeroes are assigned to R or S, either R or S will have an odd number of zeroes, and the other an even number, counting multiplicity. So case 0 is absurd: necessarily $\alpha\beta \neq 0$.

1. $\alpha\beta \neq 0$ and $\alpha\beta t - q$ is a nonzero multiple of $t - d_0$, where $d_0 = 0$, 1, or d, so that exactly one of C_0, C_1, C_d is zero. U has two simple poles on $\mathbb{C}$, at $t = d_1, d_2$ (the two singular points other than d_0); it has no other poles, and no zeroes. So each of d_1, d_2 must be a simple zero of either R or S; also, R and S can have no other simple zeroes, and no zeroes of order $k > 2$. Since by assumption the Heun equation has four singular points, each of $0, 1, d$ must be a singular point, so the coefficient of dy/dt in (3.3) must have a pole at $t = d_0$, which implies that R or S must have a zero at d_0 of the only remaining type: a double zero.

2. $\alpha\beta \neq 0$ but $\alpha\beta t - q$ is not a multiple of t, $t-1$, or $t-d$, so that none of C_0, C_1, C_d is zero. U has three poles on $\mathbb{C}$, and exactly one zero, at $t = p \equiv q/\alpha\beta$, which is simple. So each of $0, 1, d$ must be a simple zero of either R or S, and $q/\alpha\beta$ must be a triple zero of either R or S. Also, R and S can have no other simple zeroes, and no other zeroes of order $k > 2$.

In cases 1 and 2, what remain to be determined are the (additional) double zeroes of R and S, if any. That is, it must be determined whether any ordinary point of the Heun equation can be mapped to $z = 0$ or $z = 1$ with double multiplicity. But by Proposition 3.3, R can map an ordinary point $t = t_0$ to $z = 0$ (resp. $z = 1$) with double multiplicity only if the exponents of $z = 0$ (resp. $z = 1$) are $0, 1/2$.

Suppose this occurs. In case 1, if the exponents of $t = p = d_0$ are denoted $0, \gamma_0$, the exponents of $R(p)$ will be $0, \gamma_0/2$, since $t = p$ will be mapped with double multiplicity to $z = R(p)$. So if $R(t_0) = R(p)$ then γ_0 must equal 1, which, since $q = \alpha\beta d_0$, is ruled out by the assumption that each of $0, 1, d$, including d_0, is a genuine singular point. It follows that in case 1, $R(t_0) \neq R(p)$. A related argument applies in case 2. In case 2, the point p is an ordinary point of the Heun equation, and a double critical point of the $t \mapsto z$ map. So as a singular point of the hypergeometric equation, $R(p)$ must have exponents $0, 1/3$. It follows that $R(t_0) = R(p)$ is impossible.

The 'only if' half of the proposition has now been proved; the 'if' half remains. Just as the equality (3.4) implies the stated conditions on R, so the stated conditions must be shown to imply the equality (3.4). But the conditions on R are equivalent to the left-hand side and right-hand side having the same poles and zeroes, i.e., to their being the same up to a constant factor. To show the constant is unity, it is enough to consider the limit $t \to \infty$. If $\deg R = n$, then $\dot{R}/R \sim n/t$ and $\dot{S}/S \sim -n/t$, so U, i.e., the left-hand side, has asymptotic behavior $-n^2/t^2$. This will be the same as that of the right-hand side if $(\alpha\beta)/(ab) = n^2$. But $a = \alpha/n$ and $b = \beta/n$ follow from the assumption that R maps exponents to exponents. □

Proof (of Theorem 3.1). By Proposition 3.6, the preimages of $z = 0, 1$ under the polynomial R must include $t = 0, 1, d$, and in case 2 of the proposition, $t = p \equiv q/\alpha\beta$. They may also include l (additional) double zeroes of R or of S, which will be denoted $t = a_1, \dots, a_l$. Cases 1 and 2 of the theorem correspond to cases 1 and 2 of the proposition, and the subcases of the theorem correspond to distinct choices of l.

Necessarily $\deg R = |R^{-1}(0)| = |R^{-1}(1)|$, where the inverse images are defined as multisets rather than sets, to take multiplicity into account. This places tight constraints on l, since each of $0, 1, d$ (and p, in case 2) may be assigned to either $R^{-1}(0)$ or $R^{-1}(1)$, but by the proposition, all of $a_1, \dots, a_l$ must be assigned, twice, to one or the other. In case 1, one of $0, 1, d$ (denoted d_0 in the proposition) has multiplicity 2, and the other two (denoted d_1, d_2) have multiplicity 1. It follows that $0 \leq l \leq 2$, with $\deg R = l + 2$. In case 2, each of $0, 1, d$ has multiplicity 1, and p has multiplicity 3. It follows that $0 \leq l \leq 3$, with $\deg R = l + 3$. Subcases are as follows.

1(a). Case 1, $l = 0$, $\deg R = 2$. Necessarily $R^{-1}(0)$ and $R^{-1}(1)$ are $\{d_0, d_0\}$ and $\{d_1, d_2\}$, or vice versa. Without loss of generality (WLOG), assume the latter, and assume $d_0 = 1$. Then $R^{-1}(0) = \{0, d\}$ and $R^{-1}(1) = \{1, 1\}$, i.e., $S^{-1}(0) = \{1, 1\}$ and $S^{-1}(1) = \{0, d\}$. Since $t = 1$ is a double zero of S, $S(t) = C(t-1)^2$ for some C. But $S(0) = 1$, which implies $C = 1$, and $S(d) = 1$, which implies $d = 2$. So $S(t) = (t-1)^2$ and $R(t) = t(2-t)$. Since $t = 0, d$ are both mapped singly to the singular point $z = 0$ by R, their exponents must be the same as those of $z = 0$, and hence must be identical.

1(b). Case 1, $l = 1$, $\deg R = 3$. Necessarily $R^{-1}(0)$ and $R^{-1}(1)$ are $\{d_0, d_0, d_1\}$ and $\{d_2, a_1, a_1\}$, or vice versa. WLOG, assume the former, and also assume d_0, d_1, d_2 equal $1, d, 0$, respectively. Then $R^{-1}(0) = \{1, 1, d\}$ and $R^{-1}(1) = \{0, a_1, a_1\}$. It follows that $R(t) = (t-1)^2(1 - t/d)$, where d is determined by the condition that the critical point of R other than $t = 1$ (i.e., $t = a_1$) be mapped to 1. Solving $\dot{R}(t) = 0$ for $t = a_1$ yields $a_1 = (2d+1)/3$, and substitution into $R(a_1) - 1 = 0$ yields $d = 4$ or $-1/2$. But the latter is ruled out by the fact that it would imply $a_1 = 0$, which is impossible. So $d = 4$ and $a_1 = 3$. Since $t = 1, d$ are mapped to the singular point $z = 0$, doubly and singly respectively, the exponents of $t = 1$ must be twice those of $z = 0$, and the exponents of $t = d$ must be the same as those of $z = 0$.

1(c). Case 1, $l = 2$, $\deg R = 4$. Necessarily $R^{-1}(0)$ and $R^{-1}(1)$ are $\{d_0, d_0, d_1, d_2\}$ and $\{a_1, a_1, a_2, a_2\}$, or vice versa. WLOG, assume the latter, and assume $d_0 = 1$. Then $R^{-1}(0) = \{a_1, a_1, a_2, a_2\}$ and $R^{-1}(1) = \{0, 1, 1, d\}$, i.e., $S^{-1}(0) = \{0, 1, 1, d\}$ and $S^{-1}(1) = \{a_1, a_1, a_2, a_2\}$. So $S(t) = At(t-1)^2(t-d)$, where d is determined by the condition that S must have two critical points other than $t = 1$, i.e., $t = a_1, a_2$, which are mapped by S to the same critical value (in fact, to $z = 1$). Computation yields $\dot{S} = A(t-1)\left[4t^2 - (3d+2)t + d\right]$, so a_1, a_2 must be the roots of $4t^2 - (3d+2)t + d$. If the corresponding critical values are Aw_1, Aw_2, then w_1, w_2 are the roots of the polynomial in w obtained by eliminating t between $w - S(t)/A$ and $4t^2 - (3d+2)t + d$. Its discriminant turns out to be proportional to $(d-2)^2(9d^2 - 4d + 4)^3$, so the criterion for equal values is that $d = 2$ or $9d^2 - 4d + 4 = 0$. But the latter can be ruled out, since by examination it would result in a_1, a_2 being equal. So $d = 2$; $a_1, a_2 = 1 \pm \sqrt{2}/2$; and $S(t) = At(t-1)^2(t-2)$ with

$A=-4$, so that $S(a_i)=1$. Hence $R(t)=4\left[t(2-t)-\frac{1}{2}\right]^2$. Since $t=0,d$ are mapped simply to $z=1$ and $t=1$ is mapped doubly, the exponents of $t=0,d$ must be the same, and double those of $t=1$.

2(a). Case 2, $l=0$, $\deg R=3$. Necessarily $R^{-1}(0)$ and $R^{-1}(1)$ are $\{p,p,p\}$ and $\{0,1,d\}$, or vice versa. WLOG, assume the former. Then $R(t)=A(t-p)^3$ for some A. Since $t=0,1,d$ are to be mapped singly to 1, they must be the vertices of an equilateral triangle of which p is the mean, so $A=-1/p^3$ and $R=(1-t/p)^3$. WLOG, assume $d=\frac{1}{2}+i\sqrt{3}/2$, in which case $p=\frac{1}{2}+i\sqrt{3}/6$. The exponents at $t=0,1,d$ must be equal to one another, since they all equal the exponents at $z=1$.

2(b). Case 2, $l=1$, $\deg R=4$. Assume WLOG that $R^{-1}(0)=\{p,p,p,d\}$ and $R^{-1}(1)=\{0,1,a_1,a_1\}$, i.e., $S^{-1}(0)=\{0,1,a_1,a_1\}$ and $S^{-1}(1)=\{p,p,p,d\}$. It follows that $R(t)=(1-t/d)(1-t/p)^3$, but to determine d and p, it is best to focus on S. Necessarily $S(t)=At(t-1)(t-a_1)^2$, and p can be a triple zero of R iff it is a double critical point of S as well as R. The condition that S have a double critical point determines a_1. $\dot S=A(t-a_1)\left[4t^2-(3+2a_1)t+a_1\right]$, so the polynomial $4t^2-(3+2a_1)t+a_1$ must have a double root. Its discriminant is $4a_1^2-4a_1+9$, which will equal zero iff $a_1=\frac{1}{2}\pm i\sqrt{2}$. The corresponding value of the double root, i.e., the mandatory value of p, is $\frac{1}{2}\pm i\sqrt{2}/4$. The requirement that S map p to 1 implies $A=1/p(p-1)(p-a_1)^2$. d is determined as the root of $R=1-S$ other than p; some computation yields $\frac{1}{2}\pm i5\sqrt{2}/4$. WLOG the '$\pm$' in the expressions for p and d can be replaced by '$+$'. Since $t=p$ is an ordinary point and R maps $t=p$ triply to $z=0$, $z=0$ must have exponents $0,1/3$. Since R maps $t=d$ simply to $z=0$, $t=d$ must also have exponents $0,1/3$. Similarly, since R maps the ordinary point $t=a_1$ doubly to $z=1$, $z=1$ must have exponents $0,1/2$; so $t=0$ and $t=1$, which are mapped simply to $z=1$, must also.

2(c). Case 2, $l=2$, $\deg R=5$. Assume WLOG that $R^{-1}(0)=\{p,p,p,0,1\}$ and $R^{-1}(1)=\{d,a_1,a_1,a_2,a_2\}$. Then $R(t)=At(t-1)(t-p)^3$, where p is determined by R having two critical points other than $t=p$, i.e., $t=a_1,a_2$, which are mapped to the same critical value (in fact, to $z=1$). $\dot R(t)=A(t-p)^2\left[5t^2-(2p+4)t+p\right]$, so a_1,a_2 must be the roots of $5t^2-(2p+4)t+p$. If the corresponding critical values are Aw_1,Aw_2, then w_1,w_2 are the roots of the polynomial in w obtained by eliminating t between $w-R(t)/A$ and $5t^2-(2p+4)t+p$. Its discriminant turns out to be proportional to $(p^2-p+4)^3(27p^2-27p+8)^2$, so the criterion for equal values is that $p^2-p+4=0$ or $27p^2-27p+8=0$. But the former can be ruled out, since by examination it would result in a_1,a_2 being equal. The latter is true iff $p=\frac{1}{2}\pm i\sqrt{15}/18$. WLOG the plus sign may be used. This yields $a_1,a_2=\frac{1}{2}\pm2\sqrt{3}/9+i\sqrt{15}/90$. From the condition $R(a_i)=1$, it follows that $A=-i2025\sqrt{15}/64$. d is determined as the root of $R(t)-1$ other than a_1,a_2; computation yields $d=\frac{1}{2}+11\sqrt{15}/90$. Since $t=p$ is an ordinary point mapped triply to $z=0$, $z=0$ must have exponents $0,1/3$. Similarly, since R maps the ordinary points $t=a_i$ to $z=1$, $z=1$ must have exponents $0,1/2$, so $t=d$, which is mapped singly to it, must also.

2(d). Case 2, $l=3$, $\deg R=6$. Necessarily $R^{-1}(0)$ and $R^{-1}(1)$ are $\{p,p,p,0,1,d\}$ and $\{a_1,a_1,a_2,a_2,a_3,a_3\}$, or vice versa. WLOG, assume the latter. Then $R(t)=A(t-a_1)^2(t-a_2)^2(t-a_3)^2$ and $S(t)=Bt(t-1)(t-d)(t-p)^3$. Since $t=p$ is a triple zero of S, $R(t)\sim1-C(t-p)^3$ for some nonzero C. So $\sqrt{R(t)}$, defined to equal $+1$ at $t=p$, will have a similar Taylor series: $\sqrt{A}(t-a_1)(t-a_2)(t-a_3)\sim1-C(t-p)^3/2$. This is possible only if a_1,a_2,a_3 are the vertices of an equilateral triangle, and p is their mean. It follows that the roots of S other than $t=p$, i.e., $t=0,1,d$, are also the vertices of an equilateral triangle centered on p. WLOG, choose $d=\frac{1}{2}+i\sqrt{3}/2$ and $p=\frac{1}{2}+i\sqrt{3}/6$. With a bit of algebra, R can be rewritten in the form given in the theorem. Since $t=p$ is an ordinary point and R maps it triply to $z=0$, $z=0$ must have exponents $0,1/3$. Since R maps $t=0,1,d$

simply to $z = 0$, $t = 0, 1, d$ must also have exponents $0, 1/3$. □

COROLLARY 3.7. *Suppose a Heun equation has four singular points and is nontrivial ($\alpha\beta \neq 0$ or $q \neq 0$). Then the only reductions of the corresponding local Heun function Hl to the hypergeometric function ${}_2F_1$ which can be performed by a rational transformation of the independent variable involve, in fact, polynomial transformations of degrees 2, 3, 4, 5, and 6. Up to the choice of d from the possible cross-ratio orbits, the following canonical reductions are the only such reductions. Here α, β, γ are free parameters.*

$$Hl\,(2,\,\alpha\beta;\,\alpha,\,\beta,\,\gamma,\,\alpha+\beta-2\gamma+1;\,t) = {}_2F_1\,(\alpha/2,\,\beta/2;\,\gamma;\,t(2-t))\,, \tag{3.5a}$$

$$Hl\left(4,\,\alpha\beta;\,\alpha,\,\beta,\,\tfrac12,\,2(\alpha+\beta)/3;\,t\right) = {}_2F_1\left(\alpha/3,\,\beta/3;\,\tfrac12;\,1-(t-1)^2(1-t/4)\right), \tag{3.5b}$$

$$Hl\,(2,\,\alpha\beta;\,\alpha,\,\beta,\,(\alpha+\beta+2)/4,\,(\alpha+\beta)/2;\,t) = {}_2F_1\left(\alpha/4,\,\beta/4;\,(\alpha+\beta+2)/4;\,1-4\big[t(2-t)-\tfrac12\big]^2\right), \tag{3.5c}$$

$$Hl\left(\tfrac12+i\sqrt3/2,\,\alpha\beta(\tfrac12+i\sqrt3/6);\,\alpha,\,\beta,\,(\alpha+\beta+1)/3,\,(\alpha+\beta+1)/3;\,t\right) = {}_2F_1\left(\alpha/3,\,\beta/3;\,(\alpha+\beta+1)/3;\,1-\big[1-t/(\tfrac12+i\sqrt3/6)\big]^3\right), \tag{3.5d}$$

$$Hl\left(\tfrac12+i5\sqrt2/4,\,\alpha(\tfrac23-\alpha)(\tfrac12+i\sqrt2/4);\,\alpha,\,\tfrac23-\alpha,\,\tfrac12,\,\tfrac12;\,t\right) = {}_2F_1\Big(\alpha/4,\,\tfrac16-\alpha/4;\,\tfrac12;\; 1-\big[1-t/(\tfrac12+i5\sqrt2/4)\big]\big[1-t/(\tfrac12+i\sqrt2/4)\big]^3\Big), \tag{3.5e}$$

$$Hl\left(\tfrac12+i11\sqrt{15}/90,\,\alpha(\tfrac56-\alpha)(\tfrac12+i\sqrt{15}/18);\,\alpha,\,\tfrac56-\alpha,\,\tfrac23,\,\tfrac23,\,t\right) = {}_2F_1\Big(\alpha/5,\,\tfrac16-\alpha/5;\,\tfrac23;\; (-i2025\sqrt{15}/64)\,t(t-1)\big[t-(\tfrac12+i\sqrt{15}/18)\big]^3\Big), \tag{3.5f}$$

$$Hl\left(\tfrac12+i\sqrt3/2,\,\alpha(1-\alpha)(\tfrac12+i\sqrt3/6);\,\alpha,\,1-\alpha,\,\tfrac23,\,\tfrac23,\,t\right) = {}_2F_1\left(\alpha/6,\,\tfrac16-\alpha/6;\,\tfrac23;\,1-4\left\{\big[1-t/(\tfrac12+i\sqrt3/6)\big]^3-\tfrac12\right\}^2\right). \tag{3.5g}$$

Each of these equalities holds in a neighborhood of $t = 0$ whenever the two sides are defined, e.g., whenever the fifth argument of Hl and the third argument of ${}_2F_1$ are not equal to a nonpositive integer.

Remarks. The equalities of Corollary 3.7 hold even if the Heun equation has fewer than four singular points, or is trivial; but in either of those cases, additional reductions are possible. For the trivial case, see §5.

The special harmonic reduction (3.5c) is composite: it can be obtained from the special case $\gamma = (\alpha+\beta+2)/4$ of the harmonic reduction (3.5a) by applying the standard quadratic hypergeometric transformation [6]

$$ {}_2F_1\big(a,\,b;\,(a+b+1)/2;\,z\big) = {}_2F_1\big(a/2,\,b/2;\,(a+b+1)/2;\,1-4(z-\tfrac12)^2\big) \tag{3.6}$$

to its right-hand side. The special equianharmonic reduction (3.5g) can be obtained in the same way from the case $\beta = 1-\alpha$ of the equianharmonic reduction (3.5d).

One might think that (3.6) could be applied to the right-hand sides of the remaining reductions in (3.5a)–(3.5g), to generate additional composite reduction formulæ. However, there are only a few cases in which it can be applied; and it is easily checked that when it can, it imposes conditions on the parameters of Hl which require that the Heun equation of which Hl is a solution have fewer than four singular points.

Proof. Hl and ${}_2F_1$ are the local solutions of their respective equations which belong to the exponent zero at $t=0$ (resp. $z=0$), and are regular and normalized to unity there. So the corollary follows from Theorem 3.1: (3.5a)–(3.5c) correspond to subcases 1(a)–1(c), and (3.5d)–(3.5g) to subcases 2(a)–2(d). In each subcase, the Gauss parameters $(a,b;c)$ of ${}_2F_1$ are computed by first calculating the exponents at $z=0,1,\infty$, in the way explained in Remark 3. In some subcases, the polynomial map supplied in Theorem 3.1 must be chosen to be S rather than R, due to the need to map $t=0$ to $z=0$, not $z=1$, so that the transformation will reduce Hl to ${}_2F_1$, rather than to another local solution of the hypergeometric equation. □

In all, there are $28 = 56/2$ reductions of Hl to ${}_2F_1$ of this non-F-homotopic sort, rather than 56, since the R-vs.-S choice mentioned in Remark 2 does not apply, as noted in the proof. Equivalently, for each subcase of Theorem 3.1, there is one reduction of this sort for each of the possible values of d.

Each reduction listed in Corollary 3.7 corresponds to choosing $d = D$. Any other d on the cross-ratio orbit of D may be chosen, but the orbit is defined by $\triangle 01d$ being one of the triangles (at most six) similar to $\triangle 01D$, i.e., being obtained from $\triangle 01D$ by a linear transformation $L_1 \in \mathcal{L}(\mathfrak{H})$. So for any corollary subcase and choice of d, the corresponding reduction is $z = L_2(R_1(L_1(t)))$, where L_1 is constrained to map $\triangle 01d$ to $\triangle 01D$, R_1 is the canonical transformation that appears in the subcase, and $L_2 \in \mathcal{L}(\mathfrak{h})$, i.e., $L_2(z) = z$ or $1-z$, must be chosen so that $t=0$ is mapped to $z=0$.

For example, in (3.5a), which corresponds to the harmonic subcase 1(a), the canonical transformation is $R_1 = t(2-t)$, and the $d=-1$ reduction of Hl to ${}_2F_1$ is

$$Hl\big(-1,\,0;\,\alpha,\,\beta,\,\gamma,\,(\alpha+\beta-\gamma+1)/2;\,t\big) = {}_2F_1\left(\alpha/2,\,\beta/2;\,(\gamma+1)/2;\,t^2\right). \tag{3.7}$$

This reduction is obtained by choosing $L_1(t) = t+1$ and $L_2(z) = 1-z$.

In applications, it is seldom the case that the four regular singular points of an equation of Heun type are located at $0, 1, d, \infty$. But Theorem 3.1 and its corollary may readily be generalized. Consider the canonical situation when three of the four have zero as a characteristic exponent, since this may always be arranged by an F-homotopic transformation. There are two situations of interest: either the singular points include the point at infinity, and each of the finite singular points has zero as a characteristic exponent, or the location of the singular points is unrestricted. The latter includes the former. They have the respective P-symbols

$$P\left\{\begin{matrix} d_1 & d_2 & d_3 & \infty & \\ 0 & 0 & 0 & \alpha & ;s \\ 1-\gamma & 1-\delta & 1-\epsilon & \beta & \end{matrix}\right\}, \tag{3.8a}$$

$$P\left\{\begin{matrix} d_1 & d_2 & d_3 & d_4 & \\ 0 & 0 & 0 & \alpha & ;s \\ 1-\gamma & 1-\delta & 1-\epsilon & \beta & \end{matrix}\right\}. \tag{3.8b}$$

In the nomenclature of Ref. [11], they are the canonical natural general-form Heun equation and the canonical general-form Heun equation. They are transformed to the Heun equation (2.1) by the linear and Möbius transformations

$$t = \frac{s-d_1}{d_2-d_1}, \qquad t = \frac{(s-d_1)(d_2-d_4)}{(d_2-d_1)(s-d_4)}, \tag{3.9}$$

respectively. Each of the P-symbols (3.8) is accompanied by an accessory parameter. The equation specified by (3.8a) can be written as

$$\frac{d^2u}{ds^2} + \left(\frac{\gamma}{s-d_1} + \frac{\delta}{s-d_2} + \frac{\epsilon}{s-d_3}\right)\frac{du}{ds} + \frac{\alpha\beta s - q'}{(s-d_1)(s-d_2)(s-d_3)}\, u = 0, \tag{3.10}$$

where q' is the accessory parameter [11]. The equation specified by (3.8b) with $d_4 \neq \infty$ can be written as

$$\begin{aligned} &\frac{d^2u}{ds^2} + \left(\frac{\gamma}{s-d_1} + \frac{\delta}{s-d_2} + \frac{\epsilon}{s-d_3} + \frac{1-\alpha-\beta}{s-d_4}\right)\frac{du}{ds} \\ &\qquad + \frac{\alpha\beta\left[\prod_{i=1}^{3}(d_4-d_i)\right]\Big/(s-d_4) - q''}{(s-d_1)(s-d_2)(s-d_3)(s-d_4)}\, u = 0 \end{aligned} \tag{3.11}$$

where q'' is the accessory parameter [11].

There is an impediment to the satisfactory generalization of Theorem 3.1 to these two equations, which is the rigorous specification of which cases should be excluded on account of their being ‘trivial’, or having fewer than four singular points. The excluded cases should really be specified not in terms of the *ad hoc* accessory parameters q' and q'', but rather in an invariant way, in terms of an accessory parameter defined so as to be invariant under linear or Möbius transformations. (See Ref. [11], Addendum, §2.2.) However, it is clear that (3.10) is trivial, i.e., can be transformed to a trivial Heun equation by a linear transformation, iff $\alpha\beta = 0$, $q' = 0$. Also, it has fewer than four singular points if $\gamma = 0$, $q' = 0$; or $\delta = 0$, $q' = \alpha\beta$; or $\epsilon = 0$, $q' = \alpha\beta d$. Likewise, it is fairly clear that (3.11) will be trivial, i.e., can be transformed to a trivial Heun equation by a Möbius transformation, iff $\alpha\beta = 0$, $q'' = 0$. The conditions on the parameters for there to be a full set of singular points are, however, more complicated.

The first generalization of Theorem 3.1 is Corollary 3.8, which follows from Theorem 3.1 by applying the linear transformation (3.9a). It mentions a polynomial transformation, which is the composition of the $s \mapsto t$ linear transformation with the $t \mapsto z$ polynomial map of Theorem 3.1. To avoid repetition, Corollary 3.8 simply cites Theorem 3.1 for the necessary and sufficient conditions on the characteristic exponent parameters and the accessory parameter.

COROLLARY 3.8. *A canonical natural general-form Heun equation of the form (3.10), which has four singular points and is nontrivial (i.e., $\alpha\beta \neq 0$ or $q' \neq 0$), can be transformed to a hypergeometric equation of the form $(\mathfrak{h})$ by a rational substitution $z = R(s)$ iff $\alpha\beta \neq 0$, R is a polynomial, and the Heun equation satisfies the following conditions.*

$\triangle d_1 d_2 d_3$ must be similar to $\triangle 01D$, with $D = 2$ or $\frac{1}{2} + i\sqrt{3}/2$, or $D = 4$, $\frac{1}{2} + i5\sqrt{2}/4$, or $\frac{1}{2} + i11\sqrt{15}/90$. That is, it must either be a degenerate triangle consisting of three equally spaced collinear points (the harmonic case), or be an equilateral triangle (the equianharmonic case), or be similar to one of three other specified triangles, of which one is degenerate and two are isosceles. The characteristic exponent parameters γ, δ, ϵ must satisfy conditions that follow from the corresponding subcases of Theorem 3.1, and the accessory parameter q' must take a value that can be computed uniquely from the parameters γ, δ, ϵ and the choice of subcase.

For example, in the harmonic case, the two endpoints of the degenerate triangle of singular points $\triangle d_1 d_2 d_3$ must have equal exponent parameters, and q' must equal $\alpha\beta$ times the intermediate point. In this case, R will in general be a quadratic polynomial. There are two possibilities: R will map the two endpoints to $z = 0$ and the intermediate point to $z = 1$, or *vice versa. If the characteristic exponents of the intermediate point are twice those of the endpoints, then R may also be quartic: the composition of either possible quadratic polynomial with a subsequent $z \mapsto 4(z-\frac{1}{2})^2$ or $z \mapsto 4z(1-z)$ map.*

In the equianharmonic case, all three exponent parameters γ, δ, ϵ must be equal, and the accessory parameter q' must equal $\alpha\beta$ times the mean of d_1, d_2, d_3. In this case, R will in general be a cubic polynomial. There are two possibilities: R will map d_1, d_2, d_3 to $z = 0$ and their mean to $z = 1$, or vice versa. If the exponent parameters γ, δ, ϵ equal $2/3$, then R may also be sextic: the composition of either possible cubic polynomial with a subsequent $z \mapsto 4(z-\frac{1}{2})^2$ or $z \mapsto 4z(1-z)$ map.

The further generalization of Theorem 3.1 is Corollary 3.9, which follows from Theorem 3.1 by applying the Möbius transformation (3.9b). It mentions a rational substitution, which is the composition of the $s \mapsto t$ Möbius transformation with the $t \mapsto z$ polynomial map of Theorem 3.1.

COROLLARY 3.9. *A canonical general-form Heun equation of the form (3.11), which has four singular points and is nontrivial (i.e., $\alpha\beta \neq 0$ or $q'' \neq 0$), can be transformed to a hypergeometric equation of the form $(\mathfrak{h})$ by a rational substitution $z = R(s)$ iff $\alpha\beta \neq 0$, and the Heun equation satisfies the following condition.*

The cross-ratio orbit of d_1, d_2, d_3, d_4 must be the same as that of $0, 1, D, \infty$, where D takes one of the five values enumerated in Theorem 3.1. That is, the cross-ratio orbit must be the harmonic orbit, the equianharmonic orbit, or one of three specified generic orbits, one real and two non-real. The characteristic exponent parameters γ, δ, ϵ must satisfy conditions that follow from the corresponding subcases of Theorem 3.1, and the accessory parameter q'' must take a value that can be computed uniquely from the parameters γ, δ, ϵ and the choice of subcase.

Example 1. Suppose d_1, d_2, d_3, d_4 form a harmonic quadruple, i.e., they can be mapped by a Möbius transformation to the vertices of a square in $\mathbb{C}$. Moreover, two of d_1, d_2, d_3 have the same characteristic exponents, and are mapped to diagonally opposite vertices of the square. That is, of the three parameters γ, δ, ϵ, the two corresponding to a diagonally opposite pair must be equal.

In this case, provided the accessory parameter takes a value that can be computed from the other parameters, a substitution R exists. In general, it will be a degree-2 rational function, the only critical points of which are the third singular point (out of d_1, d_2, d_3) and d_4. Either R will map the two distinguished singular points to $z = 1$ and the third singular point to $z = 0$, or vice versa; and d_4 to $z = \infty$. In the special case when the characteristic exponents of the third point are twice those of the two distinguished points, degree-4 rational substitutions are also possible.

Example 2. Suppose d_1, d_2, d_3, d_4 form an equianharmonic quadruple, i.e., they can be mapped by a Möbius transformation to the vertices of a regular tetrahedron in $\mathbb{CP}^1$. Moreover, d_1, d_2, d_3 have the same characteristic exponents, i.e., $\gamma = \delta = \epsilon$.

In this case, provided the accessory parameter takes a value uniquely determined by the other parameters, a substitution R exists. In general, R will be a degree-3 rational function, the only critical points of which are the mean of d_1, d_2, d_3 with respect to d_4, and d_4. Either R will map d_1, d_2, d_3 to $z = 1$ and the mean of d_1, d_2, d_3 with respect to d_4 to $z = 0$, or vice versa; and d_4 to $z = \infty$. In the special case when the characteristic exponents of each of d_1, d_2, d_3 equal $0, 1/3$, degree-6 rational substitutions are also possible.

Remark. In Example 2, the concept of the mean of three points in $\mathbb{CP}^1$ with respect to a distinct fourth point was used. A projectively invariant definition is the following. If T is a Möbius transformation that takes d_4 ($\neq d_1, d_2, d_3$) to the point at infinity, the mean of d_1, d_2, d_3 with respect to d_4 is the point that would be mapped to the mean of Td_1, Td_2, Td_3

by T.

4. The Clarkson–Olver Transformation. The transformation discovered by Clarkson and Olver [5], which stimulated these investigations, turns out to be a special case of the equianharmonic transformation discovered in §3. Their transformation was originally given in a rather complicated form, which we shall simplify.

Recall that the Weierstrass function $\wp(u) \equiv \wp(u; g_2, g_3)$ with invariants $g_2, g_3 \in \mathbb{C}$ has a double pole at $u = 0$, and satisfies

$$\begin{aligned} \wp'^2 &= 4\wp^3 - g_2\wp - g_3 \\ &= 4(\wp - e_1)(\wp - e_2)(\wp - e_3), \end{aligned} \tag{4.1}$$

where e_1, e_2, e_3, the zeroes of the defining cubic polynomial, are the finite critical values of $\wp$. By convention, the polynomial has no quadratic term (i.e., $g_1 = 0$), so

$$e_1 + e_2 + e_3 = 0. \tag{4.2}$$

In general, $\wp$ is doubly periodic on $\mathbb{C}$, with periods denoted $2\omega, 2\omega'$. So $\wp$ can be viewed as a function on the torus $\mathbb{T} \stackrel{\text{def}}{=} \mathbb{C}/\mathcal{L}$, where $\mathcal{L} = \{2n\omega + 2n'\omega' \mid n, n' \in \mathbb{Z}\}$ is the period lattice. It turns out that the half-lattice $\{0, \omega, \omega', \omega + \omega'\} + \mathcal{L}$ comprises the critical points of $\wp$. The map $\wp : \mathbb{T} \to \mathbb{CP}^1$ is a branched double cover of the Riemann sphere, but $\mathbb{T}$ is uniquely coordinatized by the value of the pair $(\wp, \wp')$. The properties of $\mathbb{T}$ are determined by the invariant $\Delta \stackrel{\text{def}}{=} g_2^3 - 27g_3^2$, which is the discriminant of the defining polynomial. That is, Δ equals $16\prod_{\langle ij \rangle}(e_i - e_j)^2$.

If $\Delta > 0$ (the so-called real rectangular case, which predominates in applications), ω and ω' can be taken to be real and imaginary, respectively. If $\Delta < 0$ (the less familiar real rhombic case), this is not possible. However, it is possible to choose them to be complex conjugates, so that the third basic critical point $\omega_2 \stackrel{\text{def}}{=} \omega + \omega'$ is real. The distinction between $\Delta > 0$ and $\Delta < 0$ is important, since tori $\mathbb{T}$ corresponding to different values of Δ are not, in general, homeomorphic as complex analytic manifolds.

Clarkson and Olver considered the Weierstrass-form Lamé equation

$$\frac{d^2\psi}{du^2} - [j(j+1)\wp(u) + r]\,\psi = 0, \tag{4.3}$$

which can be viewed as a Fuchsian equation on $\mathbb{T}$, with exactly one regular singular point (at $(\wp, \wp') = (\infty, \infty)$) and a single accessory parameter, r. [We have altered their characteristic exponent parameter σ to $-j(j+1)/36$, to agree with the literature, and have added the accessory parameter.] In particular, they considered the case $g_2 = 0$, $g_3 \neq 0$, $r = 0$. They mapped $u \in \mathbb{T}$ to $z \in \mathbb{CP}^1$ via the formal substitution

$$u = \frac{i}{(16g_3)^{1/6}} \int^{(1-z)^{1/3}} \frac{d\tau}{\sqrt{1-\tau^3}}, \tag{4.4}$$

and showed that the Lamé equation is transformed to

$$z(1-z)\frac{d^2\psi}{dz^2} + \left(\frac{1}{2} - \frac{7}{6}z\right)\frac{d\psi}{dz} + \frac{j(j+1)}{36}\,\psi = 0. \tag{4.5}$$

This is a hypergeometric equation with $(a, b; c) = (-j/6, (j+1)/6; 1/2)$. It has characteristic exponents $0, 1/2$ at $z = 0$; $0, 1/3$ at $z = 1$; and $-j/6, (j+1)/6$ at $z = \infty$.

In elliptic function theory the case $g_2 = 0$, $g_3 \neq 0$ is called the equianharmonic case, since it yields a triple of critical values e_1, e_2, e_3 that are the vertices of an equilateral triangle in $\mathbb{C}$. If, for example, g_3 is real, then $\Delta < 0$; and by convention, e_1, e_2, e_3 correspond to $\omega, \omega + \omega', \omega'$, respectively. e_1 and e_3 are complex conjugates, and e_2 is real. The triangle $\triangle 0\omega_2\omega'$ is also equilateral (see [1], §18.13). It is customary to standardize the equianharmonic case by choosing $g_3 = 1$, so that $\Delta = -27$. This is not a major matter, however: since $\wp$ satisfies $\wp(z; 0, g_3) = g_3^{1/3}\wp(zg_3^{1/6}; 0, 1)$, all nonzero g_3 are equivalent. In fact, tori $\mathbb{T}$ defined by $g_2 = 0$, $g_3 \neq 0$ are homeomorphic as complex analytic manifolds, even if they have different values of g_3, and hence Δ.

So, what Clarkson and Olver considered was the *equianharmonic* Lamé equation, the natural domain of definition of which is a torus $\mathbb{T}$ (i.e., an elliptic curve) with special symmetries. For the Lamé equation to be viewed as a Heun equation on $\mathbb{CP}^1$, it must be transformed to its algebraic form [7], by $s = \wp(u)$. The algebraic form is

$$\frac{d^2\psi}{ds^2} + \left(\frac{1/2}{s-e_1} + \frac{1/2}{s-e_2} + \frac{1/2}{s-e_3}\right)\frac{d\psi}{ds} + \frac{[-j(j+1)/4]s - r/4}{(s-e_1)(s-e_2)(s-e_3)}\,\psi = 0. \tag{4.6}$$

If $\Delta \neq 0$, the three critical values e_1, e_2, e_3 are distinct; in which case (4.6) is a special case of (3.10), the canonical natural general form of the Heun equation, with the distinct finite singular points $d_1, d_2, d_3 = e_1, e_2, e_3$. Also, $\alpha, \beta = -j/2, (j+1)/2$, $\gamma = \delta = \epsilon = 1/2$, and $q' = r/4$. It has characteristic exponents $0, 1/2$ at $s = e_1, e_2, e_3$, and $-j/2, (j+1)/2$ at $s = \infty$.

Applying Corollary 3.8 to (4.6) yields the following.

THEOREM 4.1. *The algebraic-form Lamé equation (4.6), in the equianharmonic case $g_2 = 0$, $g_3 \neq 0$, can be transformed when $j(j+1) \neq 0$ to a hypergeometric equation of the form $(\mathfrak{h})$ by a rational transformation $z = R(s)$ iff the accessory parameter r equals zero. If this is the case, R will necessarily be a cubic polynomial; both of*

$$z = 4s^3/g_3, \qquad 1 - 4s^3/g_3 \tag{4.7}$$

will work, and they are the only possibilities.

Proof. If $j(j+1) \neq 0$, (4.6) is a nontrivial Heun equation. If $g_3 \neq 0$, then $\Delta \neq 0$ and the singular points e_i of (4.6) are distinct, so the Heun equation has four singular points; by (4.1), the e_i are the cube roots of $g_3/4$, and are the vertices of an equilateral triangle. Since $\gamma = \delta = \epsilon$, the equianharmonic case of Corollary 3.8 applies, and no other.

As noted in (4.2), the sum and hence the mean of the e_i are zero. So the polynomial $4s^3/g_3$ is the cubic polynomial that maps each singular point to 1, and their mean to zero; $1 - 4s^3/g_3$ does the reverse. These are the only possibilities for the map $s \mapsto z$, since the sextic polynomials mentioned in the equianharmonic case can be employed only if γ, δ, ϵ equal $2/3$, which is not the case here. ☐

Remark. The equianharmonic case of Corollary 3.8 also applies to the equianharmonic algebraic-form Lamé equation in the case $j(j+1) = 0$, $r \neq 0$, and guarantees it cannot be transformed to the hypergeometric equation by any rational substitution. That is because in the sense defined in §2.1, this case too is nontrivial.

COROLLARY 4.2. *The Weierstrass-form Lamé equation (4.3), in the equianharmonic case $g_2 = 0$, $g_3 \neq 0$, can be transformed when $j(j+1) \neq 0$ to a hypergeometric equation of the form $(\mathfrak{h})$ by a substitution of the form $z = R(\wp(u))$, where R is rational, iff the accessory parameter r equals zero. If this is the case, the substitutions*

$$z = 4\wp(u)^3/g_3, \qquad 1 - 4\wp(u)^3/g_3 \tag{4.8}$$

will work, and they are the only such substitutions.

In fact, applying the substitution $z = 1 - 4\wp(u)^3/g_3$ to the Lamé equation (4.3) transforms it to the hypergeometric equation (4.5) derived by Clarkson and Olver, as is readily verified. The alternative substitution yields a closely related hypergeometric equation, with the singular points $z = 0$ and $z = 1$ interchanged.

From this perspective, all that remains to be checked is the validity of the original Clarkson–Olver substitution, (4.4). It contains a multivalued elliptic integral, which may be inverted, with the aid of (4.1), to yield $z = 1-4\wp(u)^3/g_3$. Since this is listed in Corollary 4.2, the Clarkson–Olver transformation fits into the framework of §3.

A natural question is whether their transformation can be generalized. Corollary 4.2 does not offer much hope, other than allowing an arbitrary nonzero value of g_3 (which may even be non-real, so that Δ may be non-real). Actually, the harmonic case as well as the equianharmonic case of Corollary 3.8 can be applied to the algebraic-form Lamé equation. One of the resulting quadratic transformations was recently discovered by Ivanov [8]. But quadratic rather than cubic changes of the independent variable, and more general transformations of the Lamé equation, will be treated elsewhere.

The most noteworthy feature of the Clarkson–Olver transformation is that it can be performed irrespective of the choice of characteristic exponent parameter j. Only the accessory parameter r is restricted. As they remark in their paper, when $j = 1, 1/2, 1/4$, or $1/10$, it is a classical result of Schwarz that all solutions of the hypergeometric equation (4.5) are necessarily algebraic ([7], §10.3). This implies that if $r = 0$, the same is true of all solutions of the algebraic Lamé equation (4.6); which had previously been proved by Baldassarri [4], using rather different techniques. But irrespective of the choice of j, the solutions of the $r = 0$ Lamé equation reduce to solutions of the hypergeometric equation. This is quite unlike the other known classes of exact solutions of the Lamé equation, which restrict j to take values in a discrete set ([10], §2.8.4). But it is typical of hypergeometric reductions of the Heun equation; of the sort that we have considered, at least. As the theorems of §3 make clear, in general it is possible to alter characteristic exponent parameters continuously, without affecting the existence of a transformation to the hypergeometric equation.

5. The Seemingly Trivial Case $\alpha\beta = 0$, $q = 0$. In the trivial case, the Heun equation ($\mathfrak{H}$) may be solved by quadratures. Its solutions are

$$u_1(t) = \text{const}, \qquad u_2(t) = \int^t \exp\left[-\int^v \left(\frac{\gamma}{w} + \frac{\delta}{w-1} + \frac{\epsilon}{w-d}\right) dw\right] dv. \tag{5.1}$$

In the trivial limit, the local Heun function $Hl(d, q; \alpha, \beta, \gamma, \delta; t)$ degenerates to the former (in fact, to unity), and the solution belonging to the exponent $1 - \gamma$ at $t = 0$, denoted $\widetilde{Hl}(d, q; \alpha, \beta, \gamma, \delta; t)$ here, to the latter. In applications, explicit solutions, if any, are what matter. It is nonetheless interesting to examine under what circumstances a trivial Heun equation can be transformed to a hypergeometric equation.

The canonical polynomial substitutions of §3, each of which maps exponents to exponents, give rise to a large number of nonpolynomial rational transformations of the trivial Heun equation to the hypergeometric equation, via composition with certain Möbius transformations. To understand why, recall that Theorem 3.1 and Corollary 3.7 characterized, up to linear automorphisms of the two equations, the possible polynomial substitutions that may be applied to a nontrivial Heun equation. If $t \mapsto R_1(t)$ denotes a canonical polynomial transformation, the full set of polynomial reductions derived from it comprises $t \mapsto L_2(R_1(L_1(t)))$, where $L_1 \in \mathcal{L}(\mathfrak{H})$ is a linear automorphism of the Heun equation, which maps $\{0, 1, d\}$ onto $\{0, 1, D\}$, and $L_2 \in \mathcal{L}(\mathfrak{h})$ is a linear automorphism of the hypergeometric equation, which maps $\{0, 1\}$ onto $\{0, 1\}$. (The only two possibilities for L_2 are $L_2(z) = z$ and $L_2(z) = 1-z$.)

In the context of *nontrivial* Heun equations, automorphisms that are not linear could not be employed; essentially because, as discussed in §2.3, moving the point at infinity will require a compensating F-homotopic transformation. But in the trivial case, no such issue arises: by Proposition 3.5, the Heun equation is transformed to a hypergeometric equation by a rational substitution of its independent variable, $z = R(t)$, iff the substitution maps exponents to exponents. And non-F-homotopic automorphisms, including Möbius transformations which are not linear, certainly preserve exponents. The following theorem is a consequence.

THEOREM 5.1. *A Heun equation of the form* $(\mathfrak{H})$, *which has four singular points and is trivial (i.e.,* $\alpha\beta = 0$ *and* $q = 0$*), can be transformed to a hypergeometric equation of the form* $(\mathfrak{h})$ *by any rational substitution of the form* $z = M_2\left(R_1(M_1(t))\right)$, *where* $z = R_1(t)$ *is a polynomial that maps* $\{0, 1, D\}$ *to* $\{0, 1\}$, *listed (along with* D*) in one of the seven subcases of Theorem 3.1,* $M_1 \in \mathcal{M}(\mathfrak{H})$, *and* $M_1 \in \mathcal{M}(\mathfrak{h})$. *That is,* M_1 *is a Möbius transformation that maps* $\{0, 1, d, \infty\}$ *onto* $\{0, 1, D, \infty\}$, *and* M_2 *is a Möbius transformation that maps* $\{0, 1, \infty\}$ *onto* $\{0, 1, \infty\}$. *The necessary conditions on characteristic exponents stated in Theorem 3.1 must be satisfied: the conditions on exponents at specified values of* t *are to be taken as applying to the exponents at their preimages under* M_1.

Remark. As in the derivation of the reductions listed in Corollary 3.7, the Gauss parameters $(a, b; c)$ of the resulting hypergeometric function are computed by first calculating the exponents at $z = 0, 1, \infty$, using the mapping of exponents to exponents.

The following example shows how such rational substitutions are constructed. In subcase 1(a) of Theorem 3.1, i.e., the harmonic case, $D = 2$ and the polynomial tranformation is $t \mapsto z = R_1(t) = t(2-t)$; the necessary condition on exponents is that $t = 0, d$ have identical exponents. Consider $d = -1$, which is on the cross-ratio orbit of D. $M_1(t) = (t-1)/t$ can be chosen; also, let $M_2(z) = 1/z$. Then the composition

$$z = R(t) \equiv M_2\left(R_1(M_1(t))\right) = t^2/(t^2-1) \tag{5.2}$$

maps $t = 0$ to $z = 0$ and $t = \infty$ to $z = 1$ (both with double multiplicity), and $t = 1, d$ to $z = \infty$. This substitution may be applied to any trivial Heun equation with $d = -1$, provided it has identical exponents at $t = 1, d$, i.e., provided $\delta = \epsilon$.

In this example, M_1, M_2 were selected with foresight, to ensure that R maps $t = 0$ to $z = 0$. This makes it possible to regard the substitution as a reduction of Hl to ${}_2F_1$, or of $\widetilde{Hl}$ to $\widetilde{{}_2F_1}$. By computation of exponents, the reduction is

$$\begin{aligned} &\widetilde{Hl}\,(-1,\, 0;\, 0,\, \beta,\, \gamma,\, (1+\beta+\gamma)/2;\, t) \\ &\qquad = (-1)^{(\gamma-1)/2}\,\widetilde{{}_2F_1}\left(0,\, (1-\beta+\gamma)/2;\, (1+\gamma)/2;\, t^2/(t^2-1)\right). \end{aligned} \tag{5.3}$$

Here $\widetilde{Hl}, \widetilde{{}_2F_1}$ appear, since the corresponding reduction of Hl to ${}_2F_1$ is trivially valid (both sides are constant functions of t, and equal unity). The normalization factor $(-1)^{(\gamma-1)/2}$ is present because by convention $\widetilde{Hl}(t) \sim t^{1-\gamma}$ and $\widetilde{{}_2F_1}(z) \sim z^{1-c}$ in a neighborhood of $t = 0$ (resp. $z = 0$), where the principal branches are meant.

Working out the number of rational substitutions $z = R(t)$ that may be applied to trivial Heun equations, where $R(\cdot)$ is of the form $M_2\left(R_1(M_1(\cdot))\right)$, is a useful exercise. There are seven subcases of Theorem 3.1, i.e., choices for the polynomial R_1. Each subcase allows d to be chosen from an orbit consisting of m cross-ratio values: $m = 3$ in the harmonic subcases 1(a) and 1(c), $m = 2$ in the equianharmonic subcases 2(a) and 2(d), and $m = 6$ in the others. In any subcase, the 4! choices for M_1 are divided equally among the m values of d, and there are also 3! choices for M_2. So each subcase yields $(4!/m)3!$ rational substitutions for each value of d, but not all are distinct.

To count *distinct* rational substitutions for each value of d, note the following. R will map $t = 0, 1, d, \infty$ to $z = 0, 1, \infty$. Each of the subcases of Theorem 3.1 has a 'signature', specifying the cardinalities of the inverse images of the points in $\{0, 1, \infty\}$. For example, case 1(a) has signature $2; 1; 1$, which means that of those three points, one has two preimages and the other two have one. (Order is irrelevant.) In all, subcases 1(a),1(b),2(b),2(c) have signature $2; 1; 1$, and the others have signature $3; 1; 0$. By inspection, the number of distinct mappings of $t = 0, 1, d, \infty$ to $z = 0, 1, \infty$ consistent with the signature $2; 1; 1$ is 36, and the number consistent with $3; 1; 0$ is 18.

Kuiken [9] supplies a useful list of the 36 rational substitutions arising from the harmonic subcase 1(a), but states incorrectly that they are the only rational substitutions that may be applied to the trivial Heun equation. Actually, subcases 1(a)–1(c) and 2(a)–2(d) give rise to $36, 36, 18; 18, 36, 36, 18$ rational substitutions, respectively. By dividing by m, it follows that for each subcase, the number of distinct rational substitutions per value of d is $12, 6, 6; 9, 6, 6, 9$. Of these, exactly one-third map $t = 0$ to $z = 0$, rather than to $z = 1$ or $z = \infty$, and consequently yield reductions of Hl to ${}_2F_1$, or of $\widetilde{Hl}$ to $\widetilde{{}_2F_1}$. So for each subcase, the number of such reductions per value of d is $4, 2, 2; 3, 2, 2, 3$.

For example, the four such reductions with $d = -1$ that arise from the harmonic subcase 1(a) are

$$\text{(5.4a)}\quad \begin{aligned}&\widetilde{Hl}\,(-1,\, 0;\, 0,\, \beta,\, \gamma,\, (1+\beta-\gamma)/2;\, t)\\ &\quad = \widetilde{{}_2F_1}\left(0,\, \beta/2;\, (1+\gamma)/2;\, t^2\right)\end{aligned}$$

$$\text{(5.4b)}\quad \begin{aligned}&\widetilde{Hl}\,(-1,\, 0;\, 0,\, \beta,\, \gamma,\, (1+\beta+\gamma)/2;\, t)\\ &\quad = (-1)^{(\gamma-1)/2}\,\widetilde{{}_2F_1}\left(0,\, (1-\beta+\gamma)/2;\, (1+\gamma)/2;\, t^2/(t^2-1)\right)\end{aligned}$$

$$\text{(5.4c)}\quad \begin{aligned}&\widetilde{Hl}\,(-1,\, 0;\, 0,\, \beta,\, 1-\beta,\, \delta;\, t)\\ &\quad = \widetilde{{}_2F_1}\left(0,\, (1-2\beta+\delta)/2;\, 1-\beta;\, 4t/(t+1)^2\right)\end{aligned}$$

$$\text{(5.4d)}\quad \begin{aligned}&\widetilde{Hl}\,(-1,\, 0;\, 0,\, \beta,\, 1-\beta,\, \delta,\, t)\\ &\quad = (-4)^{-\beta}\,\widetilde{{}_2F_1}\left(0,\, (1-\delta)/2;\, 1-\beta;\, -4t/(t-1)^2\right)\end{aligned}$$

The reduction (5.4a), which is the only one of the four in which the degree-2 rational function R is a polynomial, is simply the trivial (i.e., $\alpha = 0$) case of (3.7), a quadratic reduction that applies to nontrivial Heun equations with $d = -1$, as well. The reduction (5.4b) was derived above as (5.3), but (5.4c) and (5.4d) are new. By examination, they are related by composition with the Möbius transformation $z \mapsto z/(z-1)$, i.e., by the automorphism in $\mathcal{M}(\mathfrak{h})$ that interchanges $z = 1$ and $z = \infty$.

Remarkably, many rational reductions of trivial Heun equations to the hypergeometric equation are *not* derived from the polynomial transformations of Theorem 3.1. The following striking degree-4 transformation is an example. The function

$$Q(t) = 1 - \left(\frac{t-1-i}{t-1+i}\right)^4 = \frac{8it(t-1)(t-2)}{(t-1+i)^4} \tag{5.5}$$

maps $t = 0, 1, d \equiv 2, \infty$ to $z = 0$; and $t = 1 \pm i$ to $z = 1, \infty$ (both with quadruple multiplicity). By Proposition 3.5, a trivial Heun equation with $d = 2$ will be transformed by Q to a hypergeometric equation iff Q maps exponents to exponents. This constrains the singular points $t = 0, 1, d, \infty$ to have the same exponents; which is possible only if each has exponents $0, 1/2$, which must also be the exponents of $z = 0$. Also, since $t = 1 \pm i$ are ordinary points of the Heun equation, the exponents of the hypergeometric equation at

$z = 1, \infty$ must be $0, 1/4$. It follows that on the level of solutions, the transformation is

$$\widetilde{Hl}(2,\,0;\,0,\,\tfrac{1}{2},\,\tfrac{1}{2},\,\tfrac{1}{2};\,t) = (i/4)^{1/2}\,\widetilde{{}_2F_1}\left(0,\,\tfrac{1}{4};\,\tfrac{1}{2};\,\frac{8it(t-1)(t-2)}{(t-1+i)^4}\right), \tag{5.6}$$

where the normalization factor $(i/4)^{1/2}$ follows from the known behavior of the function values $\widetilde{Hl}(t)$ and $\widetilde{{}_2F_1}(z)$ as $t \to 0$ and $z \to 0$. Using the definitions (2.7) and (2.9), the transformation (5.6) can be rewritten as

$$\begin{aligned} &Hl(2,\,\tfrac{3}{4};\,\tfrac{1}{2},\,1,\,\tfrac{3}{2},\,\tfrac{1}{2};\,t) \\ &\quad = (1-t)^{1/2}(1-t/2)^{1/2}\,[1-t/(1-i)]^{-2}\,{}_2F_1\left(\tfrac{1}{2},\,\tfrac{3}{4};\,\tfrac{3}{2};\,\frac{8it(t-1)(t-2)}{(t-1+i)^4}\right). \end{aligned} \tag{5.7}$$

The equality (5.7) holds in a neighborhood of $t = 0$ (it should be noted that both sides are real when t is real and sufficiently small). It is a harmonic-case reduction of Hl to ${}_2F_1$, since it applies when $d = 2$, but it is not a special case of the general harmonic reduction (3.5a), even though it is much more specialized, since it has no free parameters.

The transformation (5.7) is of a more general type than has been considered so far. In the abstract language of §2.3, it specifies a homomorphism from an $\mathrm{Aut}(\mathfrak{H})$-orbit to an $\mathrm{Aut}(\mathfrak{h})$-orbit. However, it includes a linear change of the dependent variable, resembling a complicated F-homotopy, in addition to a rational change of the independent variable. It is clear that (5.7) cannot be derived from any of the canonical (polynomial) reductions of Corollary 3.7 by applying automorphisms of ($\mathfrak{H}$) and ($\mathfrak{h}$) to its left-hand and right-hand sides. Tranformations of this more general type will be classified in Part II.

REFERENCES

[1] M. ABRAMOWITZ AND I. A. STEGUN, eds., *Handbook of Mathematical Functions*, Dover, New York, 1965.

[2] F. M. ARSCOTT, *The land beyond Bessel: A survey of higher special functions*, in Ordinary and Partial Differential Equations: Proceedings of the Sixth Dundee Conference, no. 846 in Lecture Notes in Mathematics, Springer-Verlag, 1980, pp. 26–45.

[3] R. ASKEY, *A look at the Bateman project*, in The Mathematical Legacy of Wilhelm Magnus: Groups, Geometry, and Special Functions, W. Abikoff, J. S. Birman, and K. Kuiken, eds., vol. 169 of Contemporary Mathematics, American Mathematical Society, Providence, RI, 1994, pp. 29–43.

[4] F. BALDASSARRI, *On algebraic solutions of Lamé's equation*, J. Differential Equations, 41 (1981), pp. 44–58.

[5] P. A. CLARKSON AND P. J. OLVER, *Symmetry and the Chazy equation*, J. Differential Equations, 124 (1996), pp. 225–246.

[6] A. ERDÉLYI, ed., *Higher Transcendental Functions*, McGraw–Hill, New York, 1953–55.

[7] E. HILLE, *Ordinary Differential Equations in the Complex Domain*, Wiley, New York, 1976.

[8] P. IVANOV, *On Lamé's equation of a particular kind*, J. Phys. A, 34 (2001), pp. 8145–8150.

[9] K. KUIKEN, *Heun's equation and the hypergeometric equation*, SIAM J. Math. Anal., 10 (1979), pp. 655–657.

[10] J. J. MORALES RUIZ, *Differential Galois Theory and Non-Integrability of Hamiltonian Systems*, Birkäuser, Boston/Basel, 1999.

[11] A. RONVEAUX, ed., *Heun's Differential Equations*, Oxford University Press, Oxford, 1995.

[12] R. SCHÄFKE AND D. SCHMIDT, *The connection problem for general linear ordinary differential equations at two regular singular points with applications to the theory of special functions*, SIAM J. Math. Anal., 11 (1980), pp. 848–862.

[13] C. SNOW, *Hypergeometric and Legendre Functions with Applications to Integral Equations of Potential Theory*, no. 19 in Applied Mathematics Series, National Bureau of Standards, Washington, DC, second ed., 1952.