Karen Uhlenbeck and the Calculus of Variations

Simon Donaldson

In this article we discuss the work of Karen Uhlenbeck, mainly from the 1980s, focused on variational problems in differential geometry.

The calculus of variations goes back to the 18th century. In the simplest setting we have a functional

$$F(u) = \int \Phi(u, u')dx,$$

defined on functions u of one variable x. Then the condition that $\mathcal{F}$ is stationary with respect to compactly supported variations of u is a second order differential equation—the *Euler–Lagrange equation* associated to the functional. One writes

$$\delta\mathcal{F} = \int \delta u\, \tau(u)\, dx,$$

where

$$\tau(u) = \frac{\partial\Phi}{\partial u} - \frac{d}{dx}\frac{\partial\Phi}{\partial u'}. \tag{1}$$

The Euler–Lagrange equation is $\tau(u) = 0$. Similarly for vector-valued functions of a variable $x \in \mathbf{R}^n$. Depending on the context, the functions would be required to satisfy suitable boundary conditions or, as in most of this article, might be defined on a compact manifold rather than a domain in $\mathbf{R}^n$, and u might not exactly be a function but

Simon Donaldson is a permanent member of the Simons Center for Geometry and Physics at Stony Brook University and Professor at Imperial College London. His email address is `s.donaldson@imperial.ac.uk`.

Communicated by Notices *Associate Editor Chikako Mese.*

For permission to reprint this article, please contact: `reprint-permission@ams.org`.

DOI: https://doi.org/10.1090/noti1806

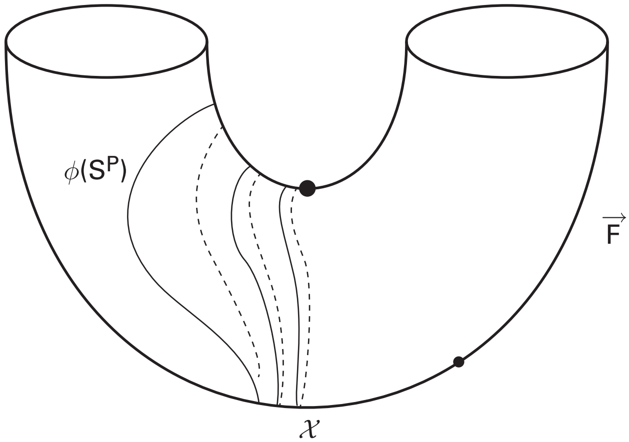


Figure 1. *Finding a critical point with a minimax sequence.*

a more complicated differential geometric object such as a map, metric, or connection. One interprets $\tau(u)$, defined as in (1), as the derivative at u of the functional $\mathcal{F}$ on a suitable infinite dimensional space $\mathcal{X}$ and the solutions of the Euler–Lagrange equation are *critical points* of $\mathcal{F}$.

A fundamental question is whether one can exploit the variational structure to establish the existence of solutions to Euler–Lagrange equations. This question came into focus at the beginning of the 20th century. Hilbert's 22nd problem from 1904 was:

Has not every regular variational problem a solution, provided certain assumptions regarding the given boundary conditions are satisfied?

If the functional $\mathcal{F}$ is bounded below one might hope to find a solution of the Euler–Lagrange equation which realises the minimum of $\mathcal{F}$ on $\mathcal{X}$. More generally, one might hope that if $\mathcal{X}$ has a complicated topology then this will force the existence of more critical points. For example, if Δ is a homotopy class of maps from the p-sphere S^p to $\mathcal{X}$ one can hope to find a critical point via a *minimax sequence*, minimising over maps $\phi \in \Delta$ the maximum of $\mathcal{F}(\phi(v))$ over points $v \in S^p$.

In Hilbert's time the only systematic results were in the case of dimension $n = 1$ and for linear problems, such as the Dirichlet problem for the Laplace equation. The development of a nonlinear theory in higher dimensions has been the scene for huge advances over the past century and provides the setting for much of Karen Uhlenbeck's work.

Harmonic Maps in Dimension 2

We begin in dimension 1 where *geodesics* in a Riemannian manifold are classical examples of solutions to a variational problem. Here we take N to be a compact, connected, Riemannian manifold and fix two points p, q in N. We take $\mathcal{X}$ to be the space of smooth paths $\gamma : [0, 1] \to N$ with $\gamma(0) = p, \gamma(1) = q$, and the energy functional

$$\mathcal{F}(\gamma) = \int_0^1 |\nabla \gamma|^2,$$

where the norm of the "velocity vector" $\nabla\gamma$ is computed using the Riemannian metric on N. The Euler–Lagrange equation is the geodesic equation, in local co-ordinates,

$$\gamma_i'' - \sum_{j,k} \Gamma^i_{jk} \gamma_j' \gamma_k' = 0,$$

where the "Christoffel symbols" Γ^i_{jk} are given by well-known formulae in terms of the metric tensor and its derivatives. In this case the variational picture works as well as one could possibly wish. There is a geodesic from p to q minimising the energy. More generally one can use minimax arguments and (at least if p and q are taken in general position) the *Morse theory* asserts that the homology of the path space $\mathcal{X}$ can be computed from a chain complex with generators corresponding to the geodesics from p to q. This can be used in both directions: facts from algebraic topology about the homology of the path space give existence results for geodesics, and, conversely, knowledge of the geodesics can feed into algebraic topology, as in Bott's proof of his periodicity theorem.

The existence of a minimising geodesic between two points can be proved in an elementary way and the original approach of Morse avoided the infinite dimensional path space $\mathcal{X}$, working instead with finite dimensional approximations, but the infinite-dimensional picture gives the best starting point for the discussion to follow. The basic point is a compactness property: *any sequence $\gamma_1, \gamma_2, \ldots$ in $\mathcal{X}$ with bounded energy has a subsequence which converges in C^0 to some continuous path from p to q.* In fact for a path $\gamma \in \mathcal{X}$ and $0 \le t_1 < t_2 \le 1$ we have

$$d(\gamma(t_1), \gamma(t_2)) \le \int_{t_1}^{t_2} |\nabla\gamma| \le \mathcal{F}(\gamma)^{1/2}\, |t_1 - t_2|^{1/2},$$

where the last step uses the Cauchy–Schwartz inequality. Thus a bound on the energy gives a $\frac{1}{2}$-Hölder bound on γ and the compactness property follows from the Ascoli-Arzela theorem.

In the same vein as the compactness principle, one can extend the energy functional $\mathcal{F}$ to a completion $\overline{\mathcal{X}}$ of $\mathcal{X}$ which is an infinite dimensional Hilbert manifold, and elements of $\overline{\mathcal{X}}$ are still continuous (in fact $\frac{1}{2}$-Hölder continuous) paths in N. In this abstract setting, Palais and Smale introduced a general "Condition C" for functionals on Hilbert manifolds, which yields a straightforward variational theory. (This was extended to Banach manifolds in early work of Uhlenbeck [24].) The drawback is that, beyond the geodesic equations, most problems of interest in differential geometry do not satisfy this Palais–Smale condition, as illustrated by the case of harmonic maps.

The harmonic map equations were first studied systematically by Eells and Sampson [5]. We now take M, N to be a pair of Riemannian manifolds (say compact) and $\mathcal{X} = \mathrm{Maps}(M, N)$ the space of smooth maps. The energy of a map $u : M \to N$ is given by the same formula

$$\mathcal{F}(u) = \int_M |\nabla u|^2,$$

where at each point $x \in M$ the quantity $|\nabla u|$ is the standard norm defined by the metrics on TM_x and $TN_{u(x)}$. In local co-ordinates the Euler Lagrange equations have the form

$$\Delta_M u_i - \sum_{jk} \Gamma^i_{jk} \nabla u_j \nabla u_k = 0, \qquad (2)$$

where Δ_M is the Laplacian on M. This is a quasi-linear elliptic system, with a nonlinear term which is quadratic in first derivatives. The equation is the natural common generalisation of the geodesic equation in N and the linear Laplace equation on M.

The key point now is that when $\dim M > 1$ the energy functional does *not* have the same compactness property. This is bound up with *Sobolev inequalities* and, most fundamentally, with the *scaling behaviour* of the functional. To explain, in part, the latter consider varying the metric g_M on M by a conformal factor λ. So λ is a strictly positive function on M and we have a new metric $\tilde{g}_M = \lambda^2 g_M$. Then one finds that the energy $\tilde{F}$ defined by this new metric is

$$\tilde{\mathcal{F}}(u) = \int_M \lambda^{2-n} |\nabla u|^2,$$

where $n = \dim M$. In particular if $n = 2$ we have $\tilde{\mathcal{F}} = \mathcal{F}$. Now take $M = S^2$ with its standard round metric and $\phi : S^2 \to S^2$ a Möbius map. This is a conformal map and it follows from the above that for any $u : S^2 \to N$ we have $\mathcal{F}(u \circ \phi) = \mathcal{F}(u)$. Since the space of Möbius maps is not compact we can construct a sequence of maps $u \circ \phi_i$ with the same energy but with no convergent subsequence.

We now recall the Sobolev inequalities. Let f be a smooth real valued function on $\mathbf{R}^n$, supported in the unit ball. We take polar co-ordinates (r, θ) in $\mathbf{R}^n$, with $\theta \in S^{n-1}$. For any fixed θ we have

$$f(0) = \int_{r=0}^1 \frac{\partial f}{\partial r} dr.$$

So, integrating over the sphere,

$$f(0) = \frac{1}{\omega_n} \int_{S^{n-1}} \int_{r=0}^1 \frac{\partial f}{\partial r} dr d\theta,$$

where ω_n is the volume of S^{n-1}. Since the Euclidean volume form is $d^n x = r^{n-1} dr d\theta$ we can write this as

$$f(0) = \frac{1}{\omega_n} \int_{B^n} |x|^{1-n} \frac{\partial f}{\partial r} \, d^n x.$$

The function $x \mapsto |x|^{1-n}$ is in L^q over the ball for any $q < n/n-1$. Let p be the conjugate exponent, with $p^{-1} + q^{-1} = 1$, so $p > n$. Then Hölder's inequality gives

$$|f(0)| \leq C_p \|\nabla f\|_{L^p}$$

where C_p is ω_n^{-1} times the $L^q(B^n)$ norm of $x \mapsto |x|^{1-n}$. The upshot is that for $p > n$ there is a continuous embedding of the *Sobolev space* L^p_1—obtained by completing in the norm $\|\nabla f\|_{L^p}$—into the continuous functions on the ball. In a similar fashion, if $p < n$ there is a continuous embedding $L^p_1 \to L^r$ for the exponent range $r \leq np/(n-p)$, which is bound up with the isoperimetric inequality in $\mathbf{R}^n$. The arithmetic relating the exponents and the dimension n reflects the scaling behaviour of the norms. If we define $f_\mu(x) = f(\mu x)$, for $\mu \geq 1$, then

$$\begin{aligned} \|f_\mu\|_{C^0} &= \|f\|_{C^0}, \\ \|f_\mu\|_{L^r} &= \mu^{-n/r} \|f\|_{L^r}, \\ \|f_\mu\|_{L^p_1} &= \mu^{1-n/p} \|f\|_{L^p_1}. \end{aligned}$$

It follows immediately that there can be no continuous embedding $L^p_1 \to C^0$ for $p < n$ or $L^p_1 \to L^r$ for $r > np/(n-p)$.

The salient part of this discussion for the harmonic map theory is that the embedding $L^p_1 \to C^0$ *fails* at the critical exponent $p = n$. (To see this, consider the function $\log \log r^{-1}$.) Taking $n = 2$ this means that the energy of a map from a 2-manifold does not control the continuity of the map and the whole picture in the 1-dimensional case breaks down. This was the fundamental difficulty addressed in the landmark paper [12] of Sacks and Uhlenbeck which showed that, with a deeper analysis, variational arguments can still be used to give general existence results.

Rather than working directly with minimising sequences, Sacks and Uhlenbeck introduced perturbed functionals on $\mathcal{X} = \mathrm{Maps}(M, N)$ (with M a compact 2-manifold):

$$\mathcal{F}_\alpha(u) = \int_M (1 + |\nabla u|^2)^\alpha.$$

For $\alpha > 1$ we are in the good Sobolev range, just as in the geodesic problem. Fix a connected component $\mathcal{X}_0$ of $\mathcal{X}$ (i.e. a homotopy class of maps from M to N). For $\alpha > 1$ there is a smooth map u_α realising the minimum of $\mathcal{F}_\alpha$ on $\mathcal{X}_0$. This map u_α satisfies the corresponding Euler–Lagrange equation, which is an elliptic PDE given by a variant of (2). The strategy is to study the convergence of u_α as α tends to 1. The main result can be outlined as follows. To simplify notation, we understand that α runs over a suitable sequence decreasing to 1.

- There is a finite set $S \subset M$ such that the u_α converge in C^∞ over $M \setminus S$.

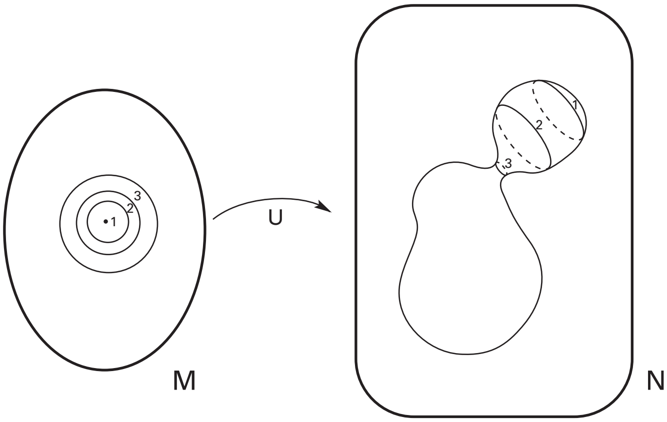


Figure 2. *Schematic representation of "bubbling."*

- The limit u of the maps u_α extends to a smooth harmonic map from M to N (which could be a constant map).
- If x is a point in S such that the u_α do *not* converge to u over a neighbourhood of x then there is a non-trivial harmonic map $v : S^2 \to N$ such that a suitable sequence of rescalings of the u_α near x converge to v.

In brief, the only way that the sequence u_α may fail to converge is by forming "bubbles," in which small discs in M are blown up into harmonic spheres in N. We illustrate the meaning of this bubbling through the example of *rational maps* of the 2-sphere. (See also the expository article [11].) For distinct points $z_1, \dots z_d$ in $\mathbf{C}$ and non-zero coefficients a_i consider the map

$$u(z) = \sum_{i=1}^{d} \frac{a_i}{z - z_i},$$

which extends to a degree d holomorphic map $u : S^2 \to S^2$ with $u(\infty) = 0$. These are in fact harmonic maps, with the same energy $8\pi d$. Take $z_1 = 0, a_1 = \epsilon$. If we make ϵ tend to 0, with the other a_i fixed, then away from 0 the maps converge to the degree $(d-1)$ map $\sum_2^d a_i(z - z_i)^{-1}$. On the other hand if we rescale about 0 by setting

$$\tilde{u}(z) = u(\epsilon z) = \frac{1}{z} + \sum_{i=2}^{d} \frac{a_i}{\epsilon z - z_i}$$

the rescaled maps converge (on compact subsets of $\mathbf{C}$) to the degree 1 map

$$v(z) = \frac{1}{z} - c$$

with $c = \sum_2^d a_i/z_i$.

A key step in the Sacks and Uhlenbeck analysis is a "small energy" statement (related to earlier results of Morrey). This says that there is some $\epsilon > 0$ such that if the energy of a map u_α on a small disc $D \subset N$ is less than ϵ then there are uniform estimates of all derivatives of u_α over the half-sized disc. The convergence result then follows from a covering argument. Roughly speaking, if the energy of the map on M is at most E then there can be at most a fixed number E/ϵ of small discs on which the map is not controlled. The crucial point is that ϵ does not depend on the size of the disc, due to the scale invariance of the energy. To sketch the proof of the small energy result, consider a simpler model equation

$$\Delta f = |\nabla f|^2, \tag{3}$$

for a function f on the unit disc in $\mathbf{C}$. Linear elliptic theory, applied to the Laplace operator, gives estimates of the schematic form

$$\|\nabla f\|_{L^q_1} \leq C\|\Delta f\|_{L^q} + \text{LOT},$$

where LOT stands for "lower order terms" in which (for this sketch) we include the fact that one will have to restrict to an interior region. Take for example $q = 4/3$. Then substituting into the equation (3) we have

$$\|\nabla f\|_{L_1^{4/3}} \leq C\||\nabla f|^2\|_{L^{4/3}} + \text{LOT} \leq C\|\nabla f\|^2_{L^{8/3}} + \text{LOT}.$$

Now in dimension 2 we have a Sobolev embedding $L_1^{4/3} \to L^4$ which yields

$$\|\nabla f\|_{L^4} \leq C\|\nabla f\|^2_{L^{8/3}} + \text{LOT}.$$

On the other hand, Hölders inequality gives the interpolation

$$\|\nabla f\|_{L^{8/3}} \leq \|\nabla f\|^{1/2}_{L^2}\|\nabla f\|^{1/2}_{L^4}.$$

So, putting everything together, one has

$$\|\nabla f\|_{L^4} \leq C\|\nabla f\|_{L^4}\|\nabla f\|_{L^2} + \text{LOT}.$$

If $\|\nabla f\|_{L^2} \leq 1/2C$ we can re-arrange this to get

$$\|\nabla f\|_{L^4} \leq \text{LOT}.$$

In other words, in the small energy regime (with $\sqrt{\epsilon} = 1/2C$) we can bootstrap using the equation to gain an estimate on a slightly stronger norm (L^4 rather than L^2) and one continues in similar fashion to get interior estimates on all higher derivatives.

This breakthrough work of Sacks and Uhlenbeck ties in with many other developments from the same era, some of which we discuss in the next section and some of which we mention briefly here.

- In *minimal submanifold theory*: when M is a 2-sphere the image of a harmonic map is a minimal surface in N (or more precisely a branched immersed submanifold). In this way, Sacks and Uhlenbeck obtained an important existence result for minimal surfaces.
- In *symplectic topology* the pseudoholomorphic curves, introduced by Gromov in 1986, are examples of harmonic maps and a variant of the Sacks–

Uhlenbeck theory is the foundation for all the ensuing developments (see, for example, [7]).

- In *PDE theory* other "critical exponent" variational problems, in which similar bubbling phenomena arise, were studied intensively (see for example the work of Brezis and Nirenberg [4]).
- In *Riemannian geometry* the Yamabe problem of finding a metric of constant scalar curvature in a given conformal class (on a manifold of dimension 3 or more) is a critical exponent variational problem for the Einstein-Hilbert functional (the integral of the scalar curvature), restricted to metrics of volume 1. Schoen proved the existence of a minimiser, completing the solution of the Yamabe problem, using a deep analysis to rule out the relevant bubbling [14].

A beautiful application of the Sacks–Uhlenbeck theory was obtained in 1988 by Micallef and Moore [8]. The argument is in the spirit of classical applications of geodesics in Riemannian geometry. Micallef and Moore considered a curvature condition on a compact Riemannian manifold N (of dimension at least 4) of having "positive curvature on isotropic 2-planes." They proved that if N satisfies this condition and is simply connected then it is a homotopy sphere (and thus, by the solution of the Poincaré conjecture, is homeomorphic to a sphere). The basic point is that a non-trivial homotopy class in $\pi_k(N)$ gives a non-trivial element of $\pi_{k-2}(\mathcal{X})$, where $\mathcal{X} = \mathrm{Maps}(S^2, N)$, which gives a starting point for a minimax argument. If N is not a homotopy sphere then by standard algebraic topology there is some k with $2 \leq k \leq \frac{1}{2}\mathrm{dim}N$ such that $\pi_k(N) \neq 0$, which implies that $\pi_{k-2}(\mathcal{X})$ is non-trivial. By developing mini-max arguments with the Sacks–Uhlenbeck theory, using the perturbed energy functional, Micallef and Moore were able to show that this leads to a non-trivial harmonic map $u : S^2 \to N$ of index at most $k-2$. (Here the index is the dimension of the space on which the second variation is strictly negative.) On the other hand the Levi–Civita connection of N defines a holomorphic structure on the pull-back $u^*(TN \otimes \mathbf{C})$ of the complexified tangent bundle. By combining results about holomorphic bundles over S^2 and a Weitzenbock formula, in which the curvature tensor of N enters, they show that the index must be at least $\frac{1}{2}\mathrm{dim}N - \frac{3}{2}$ and thus derive a contradiction.

If the sectional curvature of N is "$\frac{1}{4}$-pinched" (i.e. lies between $\frac{1}{4}$ and 1 everywhere) then N has positive curvature on isotropic 2-planes. Thus the Micallef and Moore result implies the classical sphere theorem of Berger and Klingenberg, whose proof was quite different. In turn, much more recently, Brendle and Schoen [3] proved that a (simply connected) manifold satisfying this isotropic curvature condition is in fact *diffeomorphic* to a sphere. Their proof was again quite different, using Ricci flow.

Gauge Theory in Dimension 4

From the late 1970s, mathematics was enriched by questions inspired by physics, involving *gauge fields* and the *Yang-Mills equations*. These developments were many-faceted and here we will focus on aspects related to variational theory. In this set-up one considers a fixed Riemannian manifold M and a G-bundle $P \to M$ where G is a compact Lie group. The distinctive feature, compared to most previous work in differential geometry, is that P is an auxiliary bundle not directly tied to the geometry of M. The basic objects of study are connections on P. In a local trivialisation τ of P a connection A is given by a $\mathrm{Lie}(G)$-valued 1-form A^τ. For simplicity we take G to be a matrix group, so A^τ is a matrix of 1-forms. The fundamental invariant of a connection is its curvature $F(A)$ which in the local trivialisation is given by the formula

$$F^\tau = dA^\tau + A^\tau \wedge A^\tau.$$

The Yang-Mills functional is

$$\mathcal{F}(A) = \int_M |F(A)|^2,$$

and the Euler–Lagrange equation is $d_A^* F = 0$ where d_A^* is an extension of the usual operator d^* from 2-forms to 1-forms, defined using A. This Yang-Mills equation is a nonlinear generalisation of Maxwell's equations of electromagnetism (which one obtains taking $G = U(1)$ and passing to Lorentzian signature).

In the early 1980s, Uhlenbeck proved fundamental analytical results which underpin most subsequent work in this area. The main case of interest is when the manifold M has dimension 4 and the problem is then of critical exponent type. In this dimension the Yang-Mills functional is conformally invariant and there are many analogies with the harmonic maps of surfaces discussed above. A new aspect involves gauge invariance, which does not have an analogy in the harmonic maps setting. That is, the infinite dimensional group $\mathcal{G}$ of automorphisms of the bundle P acts on the space $\mathcal{A}$ of connections, preserving the Yang-Mills functional, so the natural setting for the variational theory is the quotient space $\mathcal{A}/\mathcal{G}$. Locally we are free to change a trivialisation τ_0 by the action of a G-valued function g, which will change the local representation of the connection to

$$A^{g\tau_0} = gd(g^{-1}) + gA^{\tau_0}g^{-1}.$$

While this action of the gauge group $\mathcal{G}$ may seem unusual, within the context of PDEs, it represents a fundamental phenomenon in differential geometry. In studying Riemannian metrics, or any other kind of structure, on a manifold one has to take account of the action of the infinite-dimensional group of diffeomorphisms: for example the round metric on the sphere is only unique up to this action. Similarly, the explicit local representation of a metric depends on a choice of local co-ordinates. In fact diffeomorphism groups are much more complicated than the gauge group $\mathcal{G}$. In another direction one can have in mind the case of electromagnetism, where the connection 1-form A^τ is equivalent to the classical electric and magnetic potentials on space-time. The $\mathcal{G}$-action corresponds to the fact that these potentials are not unique.

Two papers of Uhlenbeck [25], [26] addressed both of these aspects (critical exponent and gauge choice). The paper [25] bears on the choice of an "optimal" local trivialisation τ of the bundle over a ball $B \subset M$ given a connection A. The criterion that Uhlenbeck considers is the Coulomb gauge fixing condition: $d^* A^\tau = 0$, supplemented with the boundary condition that the pairing of A^τ with the normal vector vanishes. Taking $\tau = g\tau_0$, for some arbitrary trivialisation τ_0, this becomes an equation for the G-valued function g which is a variant of the harmonic map equation, with Neumann boundary conditions. In fact the equation is the Euler–Lagrange equation associated to the functional $\|A^\tau\|_{L^2}$, on local trivialisations τ. The Yang-Mills equations in such a Coulomb gauge form an elliptic system. (Following the remarks in the previous paragraph; an analogous discussion for Riemannian metrics involves harmonic local co-ordinates, in which the Einstein equations, for example, form an elliptic system.)

The result proved by Uhlenbeck in [25] is of "small energy" type. Specialising to dimension 4 for simplicity, she shows that there is an $\epsilon > 0$ and a constant C such that if $\|F\|_{L^2(B)} < \epsilon$ there is a Coulomb gauge τ over B in which

$$\|\nabla A^\tau\|_{L^2} + \|A^\tau\|_{L^4} \le C\|F\|_{L^2}.$$

The strategy of proof uses the continuity method, applied to the family of connections given by restricting to smaller balls with the same centre, and the key point is to obtain *a priori* estimates in this family. The PDE arguments deriving these estimates have some similarity with those sketched in Section "Harmonic maps in dimension 2" above. An important subtlety arises from the critical nature of the Sobolev exponents involved. If $\tau = g\tau_0$ then an L^2 bound on ∇A^τ gives an L^2 bound on the second derivative of g but in dimension 4 this is the borderline exponent where we do not get control over the continuity of g. That makes the nonlinear operations such as $g \mapsto g^{-1}$ problematic. Uhlenbeck overcomes this problem by working with L^p for $p > 2$ and using a limiting argument.

In the companion paper [26], Uhlenbeck proves a renowned "removal of singularities" result. The statement is that a solution A of the Yang-Mills equations over the punctured ball $B^4 \setminus \{0\}$ with finite energy (i.e. with curvature $F(A)$ in L^2) extends smoothly over 0 in a suitable local trivialisation. One important application of this is that finite-energy Yang-Mills connections over $\mathbf{R}^4$ extend to the conformal compactification S^4. We will only attempt to give the flavour of the proof. Given our finite-energy solution A over the punctured ball let

$$f(r) = \int_{|x|<r} |F(A)|^2,$$

for $r < 1$. Then the derivative is

$$\frac{df}{dr} = \int_{|x|=r} |F(A)|^2.$$

The strategy is to express $f(r)$ also as a boundary integral, plus lower order terms. To give a hint of this, consider the case of an abelian group $G = U(1)$, so the connection form A^τ is an ordinary 1-form, the curvature is simply $F = dA^\tau$, and the Yang-Mills equation is $d^*F = 0$. Fix small $\epsilon < r$ and work on the annular region W where $\epsilon < |x| < r$. We can integrate by parts to write

$$\int_W |F|^2 = \int_W \langle dA^\tau, F\rangle = \int_W \langle A^\tau, d^*F\rangle + \int_{\partial W} A^\tau \wedge *F.$$

Since $d^*F = 0$ the first term on the right hand side vanishes. If one can show that the contribution from the inner boundary $|x| = \epsilon$ tends to 0 with ϵ then one concludes that

$$f(r) = \int_{|x|=r} A^\tau \wedge *F.$$

In the nonabelian case the same discussion applies up to the addition of lower-order terms, involving $A^\tau \wedge A^\tau$. The strategy is then to obtain a differential inequality of the shape

$$f(r) \le \tfrac{1}{4} r \frac{df}{dr} + \mathrm{LOT}, \tag{4}$$

by comparing the boundary terms over the 3-sphere. This differential inequality integrates to give $f(r) \le Cr^4$ and from there it is relatively straightforward to obtain an L^∞ bound on the curvature and to see that the connection can be extended over 0. The factor $\frac{1}{4}$ in (4) is obtained from an inequality over the 3-sphere. That is, any closed 2-form ω on S^3 can be expressed as $\omega = da$ where

$$\|a\|^2_{L^2(S^3)} \le \tfrac{1}{4}\|\omega\|^2_{L^2(S^3)}.$$

The main work in implementing this strategy is to construct suitable gauges over annuli in which the lower order, nonlinear terms $A^\tau \wedge A^\tau$ are controlled.

These results of Uhlenbeck lead to a Yang-Mills analogue of the Sacks–Uhlenbeck picture discussed in the previous section. This was not developed explicitly in Uhlenbeck's 1983 papers [25], [26] but results along those lines were obtained by her doctoral student S. Sedlacek [16]. Let c

be the infimum of the Yang-Mills functional on connections on $P \to X$, where X is a compact 4-manifold. Let A_i be a minimising sequence. Then there is a (possibly different) G-bundle $\tilde{P} \to X$, a Yang-Mills connection A_∞ on $\tilde{P}$, and a finite set $S \subset X$ such that, after perhaps passing to a subsequence i', the $A_{i'}$ converge to A_∞ over $X \setminus S$. (More precisely, this convergence is in $L^2_{1,\mathrm{loc}}$ and implicitly involves a sequence of bundle isomorphisms of P and $\tilde{P}$ over $X \setminus S$.) If x is a point in S such that the $A_{i'}$ do *not* converge to A_∞ over a neighbourhood of x then one obtains a non-trivial solution to the Yang-Mills equations over S^4 by a rescaling procedure similar to that in the harmonic map case. Similar statements apply to sequences of solutions to the Yang-Mills equations over X and in particular to sequences of Yang-Mills "instantons." These special solutions solve the first order equation $F = \pm * F$ and are closely analogous to the pseudoholomorphic curves in the harmonic map setting. Uhlenbeck's analytical results underpinned the applications of instanton moduli spaces to 4-manifold topology which were developed vigorously throughout the 1980s and 1990s—just as for pseudoholomorphic curves and symplectic topology. But we will concentrate here on the variational aspects.

For simplicity fix the group $G = SU(2)$; the $SU(2)$-bundles P over X are classified by an integer $k = c_2(P)$ and for each k we have a moduli space $\mathcal{M}_k$ (possibly empty) of instantons (where the sign in $F = \pm * F$ depends on the sign of k). Recall that the natural domain for the Yang-Mills functional is the infinite-dimensional quotient space $\mathcal{X}_k = \mathcal{A}_k/\mathcal{G}_k$ of connections modulo equivalence. The moduli space $\mathcal{M}_k$ is a subset of $\mathcal{X}_k$ and (if non-empty) realises the absolute minimum of the Yang-Mills functional on $\mathcal{X}_k$. In this general setting one could, optimistically, hope for a variational theory which would relate:

(1) The topology of the ambient space $\mathcal{X}_k$,
(2) The topology of $\mathcal{M}_k$,
(3) The non-minimal critical points: i.e. the solutions of the Yang-Mills equation which are not instantons.

A serious technical complication here is that the group $\mathcal{G}_k$ does not usually act freely on $\mathcal{A}_k$, so the quotient space is not a manifold. But we will not go into that further here and just say that there are suitable homology groups $H_i(\mathcal{X}_k)$, which can be studied by standard algebraic topology techniques and which have a rich and interesting structure.

Much of the work in this area in the late 1980s was driven by two specific questions.

- The *Atiyah-Jones conjecture* [1]. They considered the manifold $M = S^4$ where (roughly speaking) the space $\mathcal{X}_k$ has the homotopy type of the degree k mapping space $\mathrm{Maps}_k(S^3, S^3)$, which is in fact independent of k. The conjecture was that the inclusion $\mathcal{M}_k \to \mathcal{X}_k$ induces an isomorphism on homology groups H_i for i in a range $i \le i(k)$, where $i(k)$ tends to infinity with k. One motivation for this idea came from results of Segal in the analogous case of rational maps [17].
- Again focusing on $M = S^4$: are there *any* non-minimal solutions of the Yang-Mills equations?

A series of papers of Taubes [20], [22] developed a variational approach to the Atiyah-Jones conjecture (and generalisations to other 4-manifolds). In [20] Taubes established a lower bound on the index of any non-minimal solution over the 4-sphere. If the problem satisfied the Palais–Smale condition this index bound would imply the Atiyah-Jones conjecture (with $i(k)$ roughly $2k$) but the whole point is that this condition is not satisfied, due to the bubbling phenomenon for mini-max sequences. Nevertheless, Taubes was able to obtain many partial results through a detailed analysis of this bubbling. The Atiyah-Jones conjecture was confirmed in 1993 by Boyer, Hurtubise, Mann, and Milgram [2] but their proof worked with geometric constructions of the instanton moduli spaces, rather than variational arguments.

The second question was answered, using variational methods, by Sibner, Sibner, and Uhlenbeck in 1989 [18], showing that indeed such solutions do exist. In their proof they considered a standard S^1 action on S^4 with fixed point set a 2-sphere, an S^1-equivariant bundle P over S^4 and S^1-invariant connections on P. This invariance forces the "bubbling points" arising in variational arguments to lie on the 2-sphere $S^2 \subset S^4$ and there is a dimensional reduction of the problem to "monopoles" in 3-dimensions which has independent interest.

A connection over $\mathbf{R}^4$ which is invariant under the action of translations in one direction can be encoded as a pair (A, ϕ) of a connection A over $\mathbf{R}^3$ and an additional *Higgs field* ϕ which is a section of the adjoint vector bundle $\mathrm{ad}P$ whose fibres are copies of $\mathrm{Lie}(G)$. The Yang-Mills functional induces a Yang-Mills-Higgs functional

$$\mathcal{F}(A, \phi) = \int_{\mathbf{R}^3} |F(A)|^2 + |\nabla_A \phi|^2$$

on these pairs over $\mathbf{R}^3$. One also fixes an asymptotic condition that $|\phi|$ tends to 1 at ∞ in $\mathbf{R}^3$. In 3 dimensions we are below the critical dimension for the functional, but the noncompactness of $\mathbf{R}^3$ prevents a straightforward verification of the Palais–Smale condition. Nonetheless, in a series of papers [19], [21] Taubes developed a far-reaching variational theory in this setting. By a detailed analysis, Taubes showed that, roughly speaking, a minimax sequence can always be chosen to have energy density concentrated in a fixed large ball in $\mathbf{R}^3$ and thus obtained the necessary convergence results. In particular, using this analysis, Taubes established the existence of non-minimal critical points for the functional $\mathcal{F}(A, \phi)$.

The critical points of the Yang-Mills-Higgs functional on $\mathbf{R}^3$ yield Yang-Mills solutions over $\mathbf{R}^4$, but these do not have finite energy. However the same ideas can be applied to the S^1-action. The quotient of $S^4 \setminus S^2$ by the S^1-action can naturally be identified with the hyperbolic 3-space H^3, and S^1-invariant connections correspond to pairs (A, ϕ) over H^3. There is a crucial parameter L in the theory which from one point of view is the weight of the S^1 action on the fibres of P over S^2. From another point of view the curvature of the hyperbolic space, after suitable normalisation, is $-L^{-2}$. The fixed set S^2 can be identified with the sphere at infinity of hyperbolic space and bubbling of connections over a point in $S^2 \subset S^4$ corresponds, in the Yang-Mills-Higgs picture, to some contribution to the energy density of (A, ϕ) moving off to the corresponding point at infinity.

The key idea of Sibner, Sibner, and Uhlenbeck was to make the parameter L very large. This means that the curvature of the hyperbolic space is very small and, on sets of fixed diameter, the hyperbolic space is well-approximated by $\mathbf{R}^3$. Then they show that Taubes' arguments on $\mathbf{R}^3$ go over to this setting and are able to produce the desired non-minimal solution of the Yang-Mills equations over S^4. Later, imposing more symmetry, other solutions were found using comparatively elementary arguments [13], but the approach of Taubes, Sibner, Sibner, and Uhlenbeck is a paradigm of the way that variational arguments can be used "beyond Palais–Smale," via a delicate analysis of the behaviour of minimax sequences.

We conclude this section with a short digression from the main theme of this article. This brings in other relations between harmonic mappings of surfaces and 4-dimensional gauge theory, and touches on another very important line of work by Karen Uhlenbeck, represented by papers such as [27], [28]. In this setting the target space N is a symmetric space and the emphasis is on explicit solutions and connections with integrable systems. There is a huge literature on this subject, stretching back to work of Calabi and Chern in the 1960s, and distantly connected with the Weierstrass representation of minimal surfaces in $\mathbf{R}^3$. From around 1980 there were many contributions from theoretical physicists and any kind of proper treatment would require a separate article, so we just include a few remarks here.

As we outlined above, the dimension reduction of Yang-Mills theory on $\mathbf{R}^4$ obtained by imposing translation-invariance in one variable leads to equations for a pair (A, ϕ) on $\mathbf{R}^3$. Now reduce further by imposing translation-invariance in two directions. More precisely, write $\mathbf{R}^4 = \mathbf{R}^2_1 \times \mathbf{R}^2_2$, fix a simply-connected domain $\Omega \subset \mathbf{R}^2_1$, and consider connections on a bundle over $\Omega \times \mathbf{R}^2_2$ which are invariant under translations in $\mathbf{R}^2_2$. These correspond to pairs (A, Φ) where A is a connection on a bundle P over Ω and Φ can be viewed as a 1-form on Ω with values in the bundle $\mathrm{ad}P$. Now $A + i\Phi$ is a connection over Ω for a bundle with structure group the complexification G^c: for example if $G = U(r)$ the complexified group is $G^c = GL(r, \mathbf{C})$. The Yang-Mills instanton equations on $\mathbf{R}^4$ imply that $A + i\Phi$ is a *flat* connection. By the fundamental property of curvature, since Ω is simply-connected, this flat connection can be trivialised. The original data (A, Φ) is encoded in the reduction of the trivial G^c-bundle to the subgroup G, which amounts to a map u from Ω to the non-compact symmetric space G^c/G. For example, when $G = U(r)$ the extra data needed to recover (A, Φ) is a Hermitian metric on the fibres of the complex vector bundle, and $GL(r, \mathbf{C})/U(r)$ is the space of Hermitian metrics on $\mathbf{C}^r$. The the remaining part of the instanton equations in four dimensions is precisely the harmonic map equation for u. This is one starting point for Hitchin's theory of "stable pairs" over compact Riemann surfaces [6].

One is more interested in harmonic maps to compact symmetric spaces and, as Uhlenbeck explained in [28], this can be achieved by a modification of the set-up above. She takes $\mathbf{R}^4$ with an indefinite quadratic form of signature $(2, 2)$ and a splitting $\mathbf{R}^4 = \mathbf{R}^2_1 \times \mathbf{R}^2_2$ into positive and negative subspaces. Then the invariant instantons correspond to harmonic maps from Ω to the compact Lie group G. Other symmetric spaces can be realised as totally geodesic submanifolds in the Lie group, for example complex Grassmann manifolds in $U(r)$, and the theory can be specialised to suit. This builds a bridge between the "integrable" nature of the 2-dimensional harmonic map equations and the Penrose-Ward twistor description of Yang-Mills instantons over $\mathbf{R}^4$, although as we have indicated above much of the work on the former predates twistor theory. In her highly influential paper [28], Uhlenbeck found an action of the loop group on the space of harmonic maps from Ω to G, introduced an integer invariant "uniton number," and obtained a complete description of all harmonic maps from the Riemann sphere to G.

Higher Dimensions

In a variational theory with a critical dimension ν certain characteristic features appear when studying questions in dimensions greater than ν. In the harmonic mapping theory, for maps $u : M \to N$, the dimension in question is $n = \dim M$ and, as we saw above, the critical dimension is $\nu = 2$. A breakthrough in the higher dimensional theory was obtained by Schoen and Uhlenbeck in [12]. Suppose for simplicity that N is isometrically embedded in some Euclidean space $\mathbf{R}^k$ and define $L^2_1(M,N)$ to be the set of L^2_1 functions on M with values in the vector space $\mathbf{R}^k$ which map to N almost everywhere on M. The energy functional $\mathcal{F}$ is defined on $L^2_1(M,N)$ and Schoen and Uhlenbeck considered an energy minimising map $u \in L^2_1(M,N)$. The main points of the theory are:

- u is smooth outside a singular set $\Sigma \subset M$ which has Hausdorff dimension at most $n-3$;
- at each point x in the singular set Σ there is a *tangent map* to u.

The second item means that there is a sequence of real numbers $\sigma_i \to 0$ such that the rescaled maps

$$u_i(\xi) = u(\exp_x(\sigma_i \xi))$$

converge to a map $v : \mathbf{R}^n \to N$ which is radially invariant, and hence corresponds to a map from the the sphere S^{n-1} to N. (Here $\exp_x$ is the Riemannian exponential map and we have chosen a frame to identify TM_x with $\mathbf{R}^n$.)

To relate this to the case $n = 2$ discussed above, the general picture is that a $\mathcal{F}$-minimising sequence in Maps (M,N) can be taken to converge outside a *bubbling set* of dimension at most $n-2$ and the limit extends smoothly over the $(n-2)$-dimensional part of the bubbling set. The new feature in higher dimensions is that the limit can have a singular set of codimension 3 or more.

Two fundamental facts which underpin these results are *energy monotonicity* and *ϵ-regularity*. To explain the first, consider a smooth harmonic map $U : B^n \to N$, where B^n is the unit ball in $\mathbf{R}^n$. For $r < 1$ set

$$E(r) = \frac{1}{r^{n-2}} \int_{|x|<r} |\nabla U|^2.$$

Then one has an identity, for $r_1 < r_2$:

$$E(r_2) - E(r_1) = 2 \int_{r_1<|x|<r_2} |x|^{2-n} |\nabla_r U|^2, \qquad (5)$$

where ∇_r is the radial component of the derivative. In particular, E is an increasing function of r. The point of this is that $E(r)$ is a scale-invariant quantity. If we define $U_r(x) = U(rx)$ then $E(r)$ is the energy of the map U_r on the unit ball. The monotonicity property means that U "looks better" on a small scale, in the sense of this rescaled energy. The identity (5) follows from a very general argument, applying the stationary condition to the infinitesimal variation of U given by radial dilation. (One way of expressing this is through the theory of the stress-energy tensor.) Note that equality $E(r_2) = E(r_1)$ holds if and only if U is radially-invariant in the corresponding annulus. This is what ultimately leads to the existence of radially-invariant tangent maps.

The monotonicity identity is a feature of maps from $\mathbf{R}^n$, but a similar result holds for small balls in a general Riemannian n-manifold M. For $x \in M$ and small $r > 0$ we define

$$E_x(r) = \frac{1}{r^{n-2}} \int_{B_x(r)} |\nabla U|^2,$$

where $B_x(r)$ is the r-ball about x. Then if U is a smooth harmonic map and x is fixed the function $E_x(r)$ is increasing in r, up to harmless lower-order terms.

The ϵ-regularity theorem of Schoen and Uhlenbeck states that there is an $\epsilon > 0$ such that if u is an energy minimiser then u is smooth in a neighbourhood of x if and only if $E_x(r) < \epsilon$ for some r. An easier, related result is that if u is known to be smooth then once $E_x(r) < \epsilon$ one has *a priori* estimates (depending on r) on all derivatives in the interior ball $B_x(r/2)$. The extension to general minimising maps is one of the main technical difficulties overcome by Schoen and Uhlenbeck.

We turn now to corresponding developments in gauge theory, where the critical dimension ν is 4. A prominent achievement of Uhlenbeck in this direction is her work with Yau on the existence of Hermitian-Yang-Mills connections [29]. The setting here involves a rank r holomorphic vector bundle E over a compact complex manifold M with a Kähler metric. Any choice of Hermitian metric h on the fibres of E defines a principle $U(r)$ bundle of orthonormal frames in E and a basic lemma in complex differential geometry asserts that there is a preferred connection on this bundle, compatible with the holomorphic structure. The curvature $F = F(h)$ of this connection is a bundle-valued 2-form of type $(1,1)$ with respect to the complex structure, and we write ΛF for the inner product with the $(1,1)$ form defined by the Kähler metric. Then ΛF is a section of the bundle of endomorphisms of E. The *Hermitian-Yang-Mills equation* is a constant multiple of the identity:

$$\Lambda F = \kappa 1_E$$

(where the constant κ is determined by topology). As the name suggests, these are special solutions of the Yang-Mills equations. The result proved by Uhlenbeck and Yau is that a "stable" holomorphic vector bundle admits such a Hermitian-Yang-Mills connection. Here stability is a numerical condition on holomorphic sub-bundles, or more generally sub-sheaves, of E which was introduced by algebraic geometers studying moduli theory of holomorphic bundles. The result of Uhlenbeck and Yau confirmed conjectures made a few years before by Kobayashi and Hitchin. These extend older results of Narasimhan and Seshadri, for bundles over Riemann surfaces, and fit into a large development over the past 40 years, connecting various stability conditions in algebraic geometry with differential geometry. We will not say more about this background here but focus on the proof of Uhlenbeck and Yau.

The problem is to solve the equation $\Lambda F(h) = \kappa 1_E$ for a Hermitian metric h on E. This boils down to a second order, nonlinear, partial differential equation for h.

While this problem does not fit directly into the variational framework we have emphasised in this article, the same compactness considerations apply. Uhlenbeck and Yau use a continuity method, extending to a 1-parameter family of equations for $t \in [0,1]$ which we write schematically as $\Lambda F(h_t) = K_t$, where K_t is prescribed and $K_1 = \kappa 1_E$. They set this up so that there is a solution h_0 for $t = 0$ and the set $T \subset [0,1]$ for which a solution h_t exists is open, by an application of the implicit function theorem. The essential problem is to prove that if E is a stable holomorphic bundle then T is closed, hence equal to the whole of $[0,1]$ and in particular there is a Hermitian-Yang-Mills connection h_1.

The paper of Uhlenbeck and Yau gave two independent treatments of the core problem, one emphasising complex analysis and the other gauge theory. We will concentrate here on the latter. For a sequence $t(i) \in T$ we have connections A_i defined by the hermitian metrics $h_{t(i)}$ and the question is whether one can take a limit of the A_i. The deformation of the equations by the term K_t is rather harmless here so the situation is essentially the same as if the A_i were Yang-Mills connections. In addition, an integral identity using Chern-Weil theory shows that the Yang-Mills energy $\|F(A_i)\|^2_{L^2}$ is bounded. Then Uhlenbeck and Yau introduced a small energy result, for connections over a ball $B_x(r) \subset M$. Since the critical dimension ν is 4, the relevant normalised energy in this Yang-Mills setting is

$$E_x(r) = \frac{1}{r^{n-4}} \int_{B_x(r)} |F|^2,$$

where n is the real dimension of M. If $E_x(r)$ is below a suitable threshold there are interior bounds on all derivatives of the connection, in a suitable gauge. Then the global energy bound implies that after perhaps taking a subsequence, the A_i converge outside a closed set $S \subset M$ of Hausdorff codimension at least 4. Uhlenbeck and Yau show that if the metrics $h_{t(i)}$ do *not* converge then a suitable rescaled limit produces a holomorphic subbundle of E over $M \setminus S$. A key technical step is to show that this subundle corresponds locally to a meromorphic map to a Grassmann manifold, which implies that the subbundle extends as a coherent sheaf over all of M. The differential geometric representation of the first Chern class of this subsheaf, via curvature, shows that it violates the stability hypothesis.

The higher-dimensional discussion in Yang-Mills theory follows the pattern of that for harmonic maps above. The corresponding monotonicity formula was proved by Price [10] and a treatment of the small energy result was given by Nakajima [9]. Some years later, the theory was developed much further by Tian [23], including the existence of "tangent cones" at singular points.

This whole circle of ideas and techniques involving the dimension of singular sets, monotonicity, "small energy" results, tangent cones, etc. has had a wide-ranging impact in many branches of differential geometry over the past few decades and forms the focus of much current research activity. Apart from the cases of harmonic maps and Yang-Mills fields discussed above, prominent examples are minimal submanifold theory, where many of the ideas appeared first, and the convergence theory of Riemannian metrics with Ricci curvature bounds.

References

[1] Atiyah M, Jones J, Topological aspects of Yang-Mills theory *Comm. Math. Phys.* 61 (1978) 97-118. MR503187

[2] Boyer C, Hurtubise J, Mann B, Milgram R, The topology of instanton moduli spaces. I. The Atiyah-Jones conjecture *Ann. of Math.* 137 (1993) 561-609. MR1217348

[3] Brendle S, Schoen R, Manifolds with 1/4-pinched curvature are space forms *J. Amer. Math. Soc.* 22 (2009) 287-307. MR2449060

[4] Brézis H, Nirenberg L, Positive solutions of nonlinear elliptic equations involving critical Sobolev exponents, *Comm. Pure Appl. Math.* 36 (1983) 437-477 MR709644

[5] Eells J, Sampson J, Harmonic mappings of Riemannian manifolds, *Amer. J. Math.* 86 (1964) 109-160. MR0164306

[6] Hitchin N, The self-duality equations on a Riemann surface *Proc. Lond. Math. Soc.* 55 (1987) 59-126 MR887284

[7] McDuff D, Salamon D, *J-holomorphic curves and symplectic topology* American Mathematical Society Colloquium Publications, 52 (2004) MR2045629

[8] Micallef M, Moore J, Minimal two-spheres and the topology of manifolds with positive curvature on totally isotropic two-planes *Annals of Math.* 127 (1988) 199-227 MR924677

[9] Nakajima H, Compactness of the moduli space of Yang-Mills connections in higher dimensions *J. Math. Soc. Japan* 40 (1988) 383-392 MR945342

[10] Price P, A monotonicity formula for Yang-Mills fields *Manuscripta Math.* 43 (1983) 131-166 MR707042

[11] Parker T, What is.... a bubble tree? *Notices of the Amer. Math. Soc.* Volume 50 (2003) 666-667

[12] Sacks J, Uhlenbeck K, The existence of minimal immersions of 2-spheres *Annals of Math.* Vol 113 (1981) 1-24 MR604040

[13] Sadun L, Segert J, Non-self-dual Yang-Mills connections with nonzero Chern number *Bull. Amer. Math. Soc.* 24 (1991) 163-170 MR1067574

[14] Schoen R, Conformal deformation of a Riemannian metric to constant scalar curvature *J. Differential Geom.* 20 (1984) 479-495 MR788292

[15] Schoen R, Uhlenbeck K, A regularity theory for harmonic maps *J. Differential Geometry* 17 (1982) 307-335 MR664498

[16] Sedlacek S, A direct method for minimizing the Yang-Mills functional over 4-manifolds *Comm. Math. Phys.* 86 (1982) 515-527 MR679200

[17] Segal G, The topology of spaces of rational functions *Acta Math.* 143 (1979) 39-72. MR533892

[18] Sibner L, Sibner R, Uhlenbeck K, Solutions to Yang-Mills equations that are not self-dual *Proc. Natl. Acad. Sci.* USA 86 (1989) 8610-8613 MR1023811
[19] Taubes C, The existence of a nonminimal solution to the SU(2) Yang-Mills-Higgs equations on $\mathbf{R}^3$ *I. Comm. Math. Phys.* 86 (1982) 257-298 MR676188
[20] Taubes C, Stability in Yang-Mills theories *Comm. Math. Phys.* 91 (1983) 235-263 MR723549
[21] Taubes C, Min-max theory for the Yang-Mills-Higgs equations *Comm. Math. Phys.* 97 (1985) 473-540 MR787116
[22] Taubes C, The stable topology of self-dual moduli spaces *J. Differential Geom.* 29 (1989) 163-230 MR978084
[23] Tian G, Gauge theory and calibrated geometry *Annals of Math.* 151 (2000) 193-268 MR1745014
[24] Uhlenbeck K, Morse theory on Banach manifolds *Bull. Amer. Math. Soc.* 76 (1970) 105-106 MR0253381
[25] Uhlenbeck K, Connections with L^p bounds on curvature *Commun. Math. Phys.* 83 (1982) 31-42 MR648356
[26] Uhlenbeck K, Removable singularities in Yang-Mills fields *Commun. Math. Phys.* 83 (1982) 11-29 MR648355
[27] Uhlenbeck K, Harmonic maps into Lie groups (classical solutions of the chiral model) *J. Differential Geometry* 30 (1989) 1-50 MR1001271
[28] Uhlenbeck K, On the connection between harmonic maps and the self-dual Yang-Mills and the sine-Gordon equations *J. Geometry and Physics* 8 (1992) 283-316 MR1165884
[29] Uhlenbeck K, Yau ST, On the existence of Hermitian-Yang-Mills connections in stable vector bundles *Commun. Pure Appl. Math.* Vol XXXIX (1986) S257-S293 MR861491

Simon Donaldson

Credits

Photo of Karen Uhlenbeck is by Andrea Kane, ©IAS.
Author photo is courtesy of the author.

Celebratio Mathematica: An Interview with Karen Uhlenbeck and Three New Volumes on Women Mathematicians

Sheila Newbery

Mathematical Sciences Publishers (`msp.org`) is pleased to announce the publication of three new volumes devoted to the careers of Karen Uhlenbeck, Joan Birman, and Dusa McDuff as part of its electronic archive of mathematicians of note, *Celebratio Mathematica*. The work was supported with funding from the Mathematical Sciences Research Institute (`msri.org`) and spearheads an intensive project now underway to develop the archive's holdings on the careers and accomplishments of women mathematicians. The support made it possible to undertake a number of special projects for these volumes, one of which we highlight here: a new interview with Karen Uhlenbeck by Allyn Jackson (the latter needing no introduction to readers of the *Notices*).

Interviews like this are especially valuable to students of mathematics for whom knowledge about the careers of other women scientists can be powerfully reinforcing. In publishing them, we aim to inspire student readers especially by showing them that there are diverse paths to a career in the sciences. What leads women to mathematics is perhaps not so different in essence from what makes men choose math, but the social and institutional realities of women's careers have been different enough from those of men to warrant thoughtful attention. This fact is certainly one of the subjects of the interview.

Sheila Newbery is the managing editor of Celebratio Mathematica *(celebratio.org), an online archive of mathematicians of note published by Mathematical Sciences Publishers (msp.org). Her email address is* `sheila@msp.org`.

For permission to reprint this article, please contact: `reprint-permission@ams.org`.

DOI: http://dx.doi.org/10.1090/noti1807

The story of how Uhlenbeck became a mathematician is traced with a nuance that builds on Claudia Henrion's earlier account in *Women in Mathematics: The Addition of Difference*.[1] One detail that emerges clearly is the significance of Uhlenbeck's mother's social circle. Carolyn Keskulla was a painter, and Uhlenbeck remembers vividly as a girl that her mother's painter friends and artistic interests brought her into contact with "a lot of people who did not live normal, middle-class lives." The liveliness and eccentricity of this crowd made a lasting impression. Moreover, like her own mother (Uhlenbeck's grandmother), Carolyn was a strong, intelligent, active woman, so the value for intellectual endeavor and the uses of the imagination were early and firmly established in the family circle.

Jackson also explores the relevance of talent, and frames it instructively in the context of opportunities afforded by education: high school, first, and then a high-quality public university education. While Uhlenbeck's family background emphasized intellectual attainment, her high school gave her no specific encouragement vis-à-vis mathematics. The first seeds of mathematical inspiration were sown at university. The following excerpt from Jackson's article (see pull quote, facing page) highlights that revelatory moment, which occurred during Uhlenbeck's time as an undergraduate.

[1] *Bloomington: Indiana University Press, 1997, pp. 25–46.*

When I first read these words (while preparing the manuscript for publication), I was struck by them: for many women of Uhlenbeck's generation, the question of what one was or wasn't allowed to do was never too far from lived experience. Her word choice seemed telling. Yet the emotional impact of the episode is gloriously positive: it is one of deep intellectual excitement.

Although Uhlenbeck faced certain obstacles, such as the impossibility even as a gifted student of applying to or attending either of the nearby all-male colleges (Princeton and Rutgers), she nevertheless had access to a superb university education at the University of Michigan, and that was how she gained her first significant exposure to higher mathematics. In the post-Sputnik era of the 60s, moreover, federal agencies were issuing a clarion call to talented students, encouraging them to pursue a career in math and science after college. That posture had clear benefits for Uhlenbeck. As she puts it, "They were encouraging everybody, and women counted."

I had a very advanced course, similar to an undergraduate real analysis course. We had done limits. I went to a help session, and the teaching assistant showed how to take a derivative....It was a moment where suddenly I realized there were all sorts of things in mathematics that you "were allowed to do."
–Karen Uhlenbeck

Jackson touches on the importance of mathematical collaboration—in Uhlenbeck's case, with such colleagues as Jonathan Sacks, Lesley and Robert Sibner, S.-T. Yau, Richard Schoen, and Chuu-Lian Terng. The separate, detailed accounts here of Uhlenbeck's friendships with Yau and Lesley Sibner, in particular, vividly underscore the unpredictable swerves of social and intellectual opportunity that make up a career.

Jackson's keen sense of timing is one of her gifts as an interviewer, yet equally important is her ability to layer questions that illuminate the penumbra of intuition and wonder that motivate mathematical inquiry in the first place. So a question about mathematical "tastes" ("What kinds of mathematical problems appeal to you?") leads to a fascinating exchange about what Uhlenbeck refers to as one of the mysteries of mathematics: "[W]hy KdV comes up all over the place, in all sorts of geometric and physical problems."

In addition to Jackson's in-depth interview, we want to call readers' attention to several other unique contributions to these three new volumes:

- Cliff Taubes on "Karen Uhlenbeck's contributions to gauge theoretic analysis";
- Leonid Polterovich and Felix Schlenk's article on Dusa McDuff's contributions to "Symplectic embedding problems";
- Bill Menasco's essay "My work with Joan Birman";
- Dan Margalit and Rebecca R. Winarski's overview of Joan Birman and Hugh Hilden's collaboration in "The Birman–Hilden theory."

Mathematical Sciences Publishers will continue to build its archive: soon to be published are volumes on the careers of Mary Ellen Rudin and Cathleen Morawetz. We welcome suggestions from the mathematical community for future volumes on women mathematicians, and we are grateful to MSRI for its support in making our work available in perpetuity to a broad readership.

Sheila Newbery

Credits

Author photo is courtesy of Sheila Newbery.

Gertrude M. Cox and Statistical Design

Figure 1. Gertrude Cox in office.

Sharon L. Lohr

Introduction

On December 2, 1959, Gertrude Cox (Figure 1), Director of the Institute of Statistics at the consolidated University of North Carolina, responded to a query from a young woman named Pat Barber about career opportunities in statistics for women. Cox replied that the "field of statistics is certainly wide open to women" and described some of her own experiences as a statistician:

> In this area of experimental statistics, we cooperate with the research workers in other science areas with the planning and then with the evaluation and interpretation of their research results. I could give a list of a variety of interesting areas in which I have cooperated such as, the best methods of raising flowers in a greenhouse, development and selection of new varieties of corn, the nutritional problems among the Indian children in Guatemala, how to sample

Sharon Lohr is an Emerita Professor of Statistics at Arizona State University. Her email address is `sharon.lohr@asu.edu`.

Communicated by Notices *Associate Editor Noah Simon.*

DOI: http://dx.doi.org/10.1090/noti1811

> gold in South Africa, variations in ways to make instant frosting for cakes, how to evaluate the effectiveness of fly sprays, and many others [17].

Cox's letter reflected her view of the statistician as a partner in science, a view that, in part because of her influence, is now standard in the discipline. Her pioneering contributions and example widened opportunities in statistics around the world. To list just a few of her accomplishments, Cox:

- Founded one of the world's first statistics departments (1941) at North Carolina State College.
- Became the first woman elected to membership in the International Statistical Institute (1949) and one of the first statisticians elected to the National Academy of Sciences (1975).
- Received the O. Max Gardner Award (1959) from the University of North Carolina for "contribution to the welfare of the human race."
- Served as president of the American Statistical Association (1956) and the International Biometric Society (1968), and was founding editor of the journal *Biometrics* (1947–1955).
- Co-authored one of the most influential statistics books ever written, *Experimental Designs*, first published in 1950 and still in print.
- Championed the use of electronic computers for statistical work.

Her collaborator William G. Cochran wrote, "I doubt if anyone contributed more than Gertrude Cox to building up the profession of statistics as we know it today" [11].

Early Career

Few would have predicted in 1924 that Gertrude Mary Cox would become one of the most influential statisticians of the twentieth century. She was then a 24-year-old housemother for 16 boys at a Montana orphanage, having previously taught in a one-room schoolhouse in Iowa and studied at the Iowa National Bible Training School [21].

Cox enrolled in Iowa State College to obtain the training and credentials needed for her planned career as an orphanage superintendent. She explained in a 1975 interview how she became a statistician: she took courses in math because she liked it and it was "the easiest subject," giving her time to also take the classes in psychology and crafts she would need in her chosen career. She became interested in statistics after her calculus professor, George Snedecor, invited her to work as a computer in the Mathematical Statistical Service Center. She reminisced, "As soon as I could learn to use that math knowledge with people and their orientation, it became life" [19].

Through the 1950s, a "computer" referred to a person—usually a woman—who performed calculations on a hand-operated machine such as the one seen on the table by the radiator in Figure 1. Women were hired for this work because they were thought to be more patient with the tedious calculations than men—and, incidentally, could also be paid much less [19, 20, 30].

Many of the statistical computations involved finding correlation or regression coefficients—calculations for a regression model with a large number of independent variables could take weeks [25]. The main result in Cox's first publication [13] was a table of correlations between scores on the Iowa State College Aptitude Test, high school subjects, and college courses.

Cox received her bachelor's degree in mathematics in 1929. She stayed on at Iowa State to earn the first master's degree awarded in statistics from the Department of Mathematics in 1931, with Snedecor as advisor.

In 1933, while she was midway through a doctoral program in psychology at University of California, Berkeley, Snedecor invited her to return to Iowa State, writing, "I … am rapidly being drawn into statistical responsibilities for a large part of the College. Would you like to be a part of this? I think the opportunity is great. Are you interested? Immediately you would have charge of the girls, 140 calculating machines, and all the *stray* jobs that I can rustle for you" [18].

As one of three initial faculty members in the Iowa State Statistical Laboratory, Cox supervised the computers performing data analyses. She visited laboratories and fields to see how the data were collected, which led to collaborations with researchers to develop experimental designs and analyses that would best answer the scientific questions. Her classes in experimental design attracted students from across the campus, and she soon became known as an expert in the field. Initially hired as an assistant to Snedecor, Cox was appointed Research Assistant Professor in 1939.

In 1941 Cox became the first female full professor and the first female department head at North Carolina State College, charged with developing a department that would provide statistical expertise to researchers. Her colleague Richard Anderson related how the appointment was made:

> In 1940 Snedecor was asked to recommend candidates to head the new Department of Experimental Statistics in the School of Agriculture at North Carolina State College. "Why didn't you put my name on list?" Gertrude asked when he showed her his all-male list of candidates, and her name was added to the accompanying letter in the following postscript: "If you would consider a woman for this position, I would recommend Gertrude Cox of my staff." This terse note was to have far-reaching consequences for statistics, for not only was Gertrude considered, she was selected [10].

Historian Margaret Rossiter described how unusual it was for a woman to be considered for a position as department head in the 1940s: "As for department chairmanships, the lowest level of academic administration, women scientists still so rarely held these positions at coeducational institutions in the forties, fifties, and mid-sixties that one can almost count these exceptions on two hands." Rossiter singled out Cox as the most successful of this handful, noting that she, unlike many other university women of the time, ended up getting credit for her accomplishments in building her department and scientific discipline: "She not only managed to ride the wave of Big Science in the 1950s and 1960s but to be enough ahead of it to shape the form it took and the impact it had on her university, field, and region" [29].

Designing Experiments

Cox began her new position in North Carolina with the same energy she had shown in her work at Iowa State. She immediately started establishing training programs, hiring faculty members, collaborating with scientists, promoting statistics in the university and nationally, and teaching classes on experimental design. She expanded her mimeographed notes from the design classes into the book *Experimental Designs* [12], published with collaborator William G. Cochran in 1950.

Experimental Designs emphasized three principles:

1. Statisticians need to be involved in the research from the planning stages: the first steps, setting out the objectives of the experiment and planning the analysis, are crucial. Often, one of the statistician's most valuable contributions arises "by getting the investigator to explain clearly why he is doing the experiment, to justify the experimental treatments whose effects he proposes to compare, and to defend his claim that the completed experiment will enable its objectives to be realized." When a statistician is consulted only after the data are collected and discovers that the poorly planned experiment cannot answer the research questions, "[i]n these unhappy circumstances, about all that can be done is to indicate, if possible, how to avoid this outcome in future experiments" [12, pp. 9, 10].
2. Randomize everything that can be randomized. "Randomization is somewhat analogous to insurance, in that it is a precaution against disturbances that may or may not occur and that may or may not be serious if they do occur. It is generally advisable to take the trouble to randomize even when it is not expected that there will be any serious bias from failure to randomize. The experimenter is thus protected against unusual events that upset his expectations" [12, p. 8].
3. Use blocking whenever possible to reduce the effects of variability. Blocks are homogeneous groups of experimental units: for example, identical twins, neighboring agricultural plots, batches of raw material, cancer patients with similar demographics and disease stage, schools in the same city, or experimental runs done on the same day. When treatments are randomly assigned to experimental units within blocks—one twin is randomly chosen to receive treatment A, and the other treatment B—the block-to-block variability is removed from the treatment comparison. If blocks represent a range of experimental conditions, results from the blocked experiment have wider applicability as well as increased precision for estimating treatment effects.

The 1950 book and its second edition in 1957 set out detailed plans for Latin square, factorial, fractional factorial, split plot, lattice, balanced incomplete block, and other designs. Each design description started with examples, followed by a discussion of when the design was suitable and detailed instructions for how to perform randomization. Then came one or more detailed case studies, showing why that design had been chosen for each experiment and how it had been randomized, and taking the reader step-by-step through the calculations needed to construct the analysis of variance table and estimate the standard errors for differences of treatment means. The authors also discussed how to estimate the efficiency of the design relative to a completely randomized design and how to do the calculations for the unbalanced structure that resulted when one or more experimental runs had missing data.

The chapters on the complex designs contained tables of designs for different block sizes and numbers of treatments. Today, statistical software quickly calculates optimal designs for almost any experimental structure, but in 1950 printed design tables were needed, particularly when there was more than one blocking variable or when the number of treatments (t) exceeded the number of experimental units in a block (k).

For the latter situation, a balanced incomplete block design was recommended, where each pair of treatments occurs together in the same number of blocks. Of course such a design can always be constructed by using all combinations of the t treatments taken k at a time, but *Experimental Designs* laid out the designs that met the constraint with the smallest numbers of experimental units. For example, the smallest balanced incomplete block design with seven treatments and blocks of size four required only seven blocks and twenty-eight experimental units—one-fifth the size of the fully combinatorial design.

Cox's experience as a consulting statistician can be seen on every page of the book. Her background as a computer is also apparent: each set of instructions for calculating an analysis of variance table came with practical tips and quality checks for ensuring the calculations are accurate. Indeed, the first experiment described in *Experimental Designs* compared the speed of two calculating machines, A and B, using a cross-over design, where the same person computed the sum of squares of 10 sets (blocks) of 27 numbers on each machine; machine B turned out to be significantly faster, taking only 2 minutes 13.6 seconds, on average, to calculate the sum of squares for 27 numbers.

When Cox began her career, randomization was seldom used to protect against systematic errors or to promote valid inferences from experiments; some thought that randomization conflicted with attempts to control variation [25]. She viewed randomization as the distinguishing feature of modern statistical experimental design, and the feature that allowed proper inferences to be drawn from the results.

Cox emphasized the importance of randomization for each case study in the book. In the calculating machine experiment, for example, randomizing the machine order was essential. If the sums of squares for each block of numbers were computed first on machine A and then on machine B, and machine B turned out to be faster, one could not attribute the difference to the machines; it could have occurred because the operator became familiar with the numbers after entering them on the first machine and was able to enter them more quickly on the second. By randomly assigning machine A to be first for five of the blocks and machine B to be first for the other five, Cox could separate out the order effect and conclude that the speed difference was indeed due to the machines [12].

Cox advised the statistician to "use the simplest design that meets the needs of the experiment" [14]. In many of the experiments she consulted on, the simplest design meeting cost constraints needed blocking or other types of restricted randomization, and she and her staff tailored and developed designs for each experiment. From 1942 to 1948, all but 59 of the 6,317 experiments performed at the North Carolina Agricultural Experiment Station involved some form of blocking; 62 percent were randomized complete block designs [14]. She strove to develop ways of conducting experiments "so that the greatest amount of information can be obtained with the least expenditure of time and money" [6].

Experimental Designs is still widely used by persons designing experiments. The many experimental researchers who have recently relied on the book for guidance include Wood and Porter [33], who adopted a Latin square design to study the effects of presenting factual information to persons with strong political views, and Reeves et al. [28], who used a balanced incomplete block design to compare community and hospital eye care for persons with macular degeneration.

Designing the Statistical Profession

Of equal importance to Cox's contributions in designing experiments were her contributions in shaping the discipline of statistics.

Statistical Training

One of her earliest activities in North Carolina was establishing a summer training program in statistics. In the first six-week program, during June and July of 1941, Cox taught beginning and advanced courses on design of experiments, Snedecor taught two courses on applied statistics, and Harold Hotelling taught mathematical statistics; Ronald Fisher had been scheduled to teach, but wartime authorities in London withheld authorization for his travel. During the program, all faculty and staff members were available for individual consultations with students about statistical problems [5, 3, 21]. These courses and three affiliated one-week conferences drew 243 registrants from around the country, including many who were, or were to become, leaders of the statistics profession.

Instruction was not limited to future statisticians. The department taught multiple courses to help state government workers and other persons in the community. Cox, in addition to her administrative and other teaching duties, offered an introductory course on experimental statistics intended for tabulating clerks and computers.

In 1943, Cox intensified the department's efforts to provide trained statisticians to meet wartime and postwar needs. The summer session in 1943 offered "[f]our intensive courses in applied statistics, designed to appeal to young women who are college graduates or advanced undergraduates," including training in sampling methods. Cox said, "This training is offered because of the extreme importance of having efficient workers to help with rush work. There are numerous sampling investigations now in progress, such as those for locating sources and requirements of farm labor as well as those for studying food production and distribution problems. Training in machine problems and statistical and sampling methods is of immediate value in prosecution of the war with limited manpower" [2].

By 1946, statistics in North Carolina had grown under Cox's leadership to include the Department of Experimental Statistics at North Carolina State College and the Department of Mathematical Statistics at the University of North Carolina. Both departments were incorporated in the Institute of Statistics, which Cox directed [27]. After retiring from the university in 1960, Cox led the statistical research division at the newly formed Research Triangle Institute. She continued promoting statistics after her second "retirement" in 1965, traveling around the world to provide statistical advice and help establish statistical programs. Between travels, she served on advisory boards for the US Census Bureau, the Department of Agriculture, the National Science Foundation, and many other organizations.

When Cox established the Department of Experimental Statistics at North Carolina State in 1941, there were only a handful of statistics departments in the world: the first, Karl Pearson's Department of Applied Statistics at University College London, had been established in 1911. In general, mathematical statistics classes were taught in mathematics departments; applied statistics classes were taught in a department of agriculture, psychology, biology, or another discipline. In each, statistics was viewed as a subfield of the discipline where it was taught. Harold Hotelling, who later joined Cox's Institute of Statistics, wrote in 1940 that a great deal of the current knowledge in statistics was still in the form of oral tradition and "the seeker after truth regarding statistical theory must make his way through or around an enormous amount of trash and downright error. The great accumulation of published writings on statistical theory and methods by authors who have not sufficiently studied the subject is even more dangerous than the classroom teaching by the same people" [23].

Cox insisted that students receive a thorough grounding in mathematical theory and applications of statistics, and that they gain experience in collaborating with scientists. The universities and organizations that consulted her about establishing statistics programs inherited this philosophy, and most statistics departments today are organized around the principles she advocated for training students.

The American Statistical Association's recent guidelines for graduate and undergraduate programs in statistics [8, 9], urging that "graduates should have a solid foundation in statistical theory and methods" as well as experience with collaborating on real problems and designing studies, repeat many of the principles for statistical education that Cox outlined in 1953:

> It is the statistician's duty to keep informed of the rapidly expanding knowledge of statistics and to make such information available to the users of statistics. This combination of a thorough knowledge of statistical theory and method along with adequate competence in the field of application requires that the consultant statistician be a person of substantial ability.... A close integration between theory and applications constitutes the best foundation for important advances in the science of statistics [15].

Statistics as a Collaborative Discipline

Cox held that statistics is by its nature collaborative. Although statisticians engage in a wide range of theoretical and applied investigations, those investigations need to be directed toward "solving problems concerned with decision making" [16]. "The cooperation involved when the statistician consults and works with researchers in other fields is an advantage to all. Also, the consulting or applied statistician in his daily use of statistics encounters new problems which call for help from the theoretical statistician. The theoretical statistician requires the stimulus of practical needs which lead him into useful developments of new techniques" [18].

Equally important was the presentation of results, and a good experimental design leads to clear findings: "close cooperation between the research worker and the statistician before the experiment is started—planning the experiment so that the statistics collected will be easily interpreted by the average reasonably intelligent person" [1].

Cox practiced what she preached. Less than a week after her arrival at North Carolina State College in 1940, "she was out trooping over a soybean field near Raleigh, helping an Experiment Station agronomist work out the best set-up for an experiment. She has also visited several of the test farms for the same purpose" [1]. Throughout her career, she tirelessly promoted statistics around the world, providing expertise and helping develop programs in statistics. Her travels included consultations in Egypt, Thailand, South Africa, Guatemala, Japan, Hong Kong, Lebanon, Malaysia, Brazil, and Honduras.

Cox encouraged collaboration and sound statistical practice in her many presentations at US and international statistical conferences. She also viewed the community as a partner in statistical activity and regularly spoke about statistics and her travels to civic organizations and women's clubs in Raleigh. The *Raleigh News and Observer* reported on many of Cox's local talks. In 1954, for example, they wrote: "A talk by Gertrude Cox, director of Statistics at the University of North Carolina, was a highlight of last night's meeting of the Lewis school PTA." Her talk was followed by a presentation of a minuet from Mrs. Hicks's fourth grade class [4].

Cox provided statistical expertise locally, as well as internationally, throughout her years in Raleigh. In 1975, for example, she was asked to evaluate a controversial statistical investigation on the effectiveness of kindergarten in North Carolina. The investigators had selected 18 schools for the assessment but had omitted one school—whose results would have changed the conclusions—from the analysis. Cox's primary recommendation was that analysts should not be selective in choosing data to be analyzed unless there is justification, and she argued that the small sample size and possible selection biases made it difficult to draw clear conclusions from the study. She concluded that the investigators "could use a great deal more help from qualified statisticians" [7]—a gentle way of saying that the controversy could have been avoided if the investigators had consulted a statistician before conducting the study.

Computation and Statistics

Cox, well aware of the importance of computation to the field of statistics, established a computing laboratory soon after moving to North Carolina. The laboratory performed computations for statistical analyses as well as for other units on campus. During World War II, the department offered classes to train women as computers for the war effort.

Perhaps because of her work as a human computer, Cox was one of the first persons in statistics to embrace the ability of "electronic computing machines" to contribute to the discipline. She immediately saw their use for regression problems and computing standard errors for complex sampling designs, and she forecast that they would soon allow statisticians to "open up even wider frontiers" in statistics [16].

Not surprisingly, Cox's department was one of the first in the country to acquire one of the new IBM-650 electronic computing machines, in 1956 [27]. Computations for large regression models could now be done in less than

20 minutes rather than taking weeks. Some of the earliest computer programs for regression and analysis of variance were written at North Carolina State College [22].

Cox's interest in computational issues continued well after her retirement from the university. In the early 1970s, she provided expertise to the Department of Health, Education, and Welfare on statistical, computational, and privacy issues relating to the proposed use of the Social Security number as a universal personal identifier and, more generally, to the large amounts of personal data that were being collected in computer-based record-keeping systems. The 1973 report of the Advisory Committee on Automated Personal Data Systems set forth principles—the Code of Fair Information Practices—that became the foundation of subsequent US privacy legislation [32, 31]. The report's recommendations reflected Cox's strong views that an individual has a right to know how his or her data are being used.

Statistical Frontiers

Cox summarized her vision for statistics in her 1956 address as President of the American Statistical Association, titled "Statistical Frontiers." She invited the audience to tour the three major continents of the statistical universe: "(1) descriptive statistics, (2) design of experiments and investigations, and (3) analysis and theory" [16]. As she visited each continent, she briefly described some of the "well developed countries" where statisticians have developed many techniques for design and analysis, and she then gave examples of frontiers needing more exploration.

The descriptive statistics continent, although having the longest history of exploration, nevertheless had multiple frontiers. Cox noted that although statistical tabulations were common, too few persons described the variability of a population or the uncertainty of an estimate. She also emphasized the statistician's contributions to the presentation of results country, where "you will be asked to swear allegiance to logical organization, preciseness, and ease of comprehension" [16, p. 3].

The longest sojourn of the tour, not surprisingly, was in the design of experiments and investigations (sampling) continent. Cox foresaw the survey sampling research problems that would arise in future decades, such as the need for statistical methods to assess and control nonsampling errors, and she anticipated the development of computer-intensive methods for estimating variances [24].

While visiting the analysis and theory continent, Cox mentioned some of the frontiers of the late 1950s such as variance component models and nonparametric methods. She also discerned the fundamental problems of inference facing future statisticians in these general frontiers. The methods of statistical inference that work for data from a designed experiment or carefully collected probability sample do not necessarily apply to data that happen to be conveniently at hand. She wrote, "How far are we justified in using statistical methods based on probability theory for the analysis of nonexperimental data? Much of the data used in the descriptive methods continent are observational or nonexperimental records" [16].

Cox's comments are relevant to many of today's frontiers in statistics. One frontier in 2019 concerns making inferences from large observational data sets such as credit card transactions, electronic medical records, sensor data, or internet activity. Statistics from "big data" are often presented without any measures of uncertainty.

Participants in the 2017 National Academies of Sciences workshop on "Refining the Concept of Scientific Inference When Working with Big Data" echoed Cox's views on the need for statistical collaboration, carefully designed experiments, and appropriate statistical inference. In their report they wrote:

- "[T]oo often statisticians become involved in scientific research projects only after experiments have been designed and data collected. Inadequate involvement of statisticians in such 'upstream' activities can negatively impact 'downstream' inference, owing to suboptimal collection of information necessary for reliable inference" [26, p. 5].
- "[B]igger data does not necessarily lead to better inferences," in part "because a lot of big data is collected opportunistically instead of through randomized experiments or probability samples designed specifically for the inference task at hand" [26, p. 14].
- "Without careful consideration of the suitability of both available data and the statistical models applied, analysis of big data may result in misleading correlations and false discoveries, which can potentially undermine confidence in scientific research if the results are not reproducible" [26, p. 1].

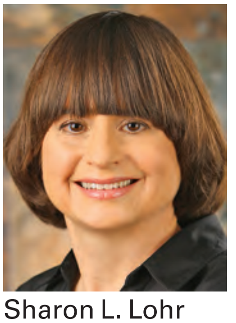
Most of Cox's views on statistics do not seem revolutionary to a statistician in 2019. That is because Cox helped define the profession of statistics from her entrance in the 1920s until her death in 1978. Her vision of the statistician as a partner in science—who collaborates on designing and analyzing studies, and who can develop new statistical theory as needed—characterizes the discipline today. She promoted sound statistical practice in the department and institutes she founded, in the community, and around the world.

As she said in 1940, "There is fascination about experimental work. In searching the unknown for new truths, there is mystery, and there is adventure, and there is the thrill of discovery" [1].

References

1. New 'statistics' about statistics: Miss Gertrude Cox sets about to correct misconceptions of her work, *Raleigh News and Observer* (December 16, 1940), 10.
2. Statistics courses to start at State, *Raleigh News and Observer* (June 1, 1943), 14.
3. Summer session beginning today, *Raleigh News and Observer* (June 16, 1941), 14.
4. Statistics specialist heard at PTA meet, *Raleigh News and Observer* (March 10, 1954), 9.
5. Statisticians to have summer course here, *Raleigh News and Observer* (March 17, 1941), 12.
6. Nutrition study to open Monday, *Raleigh News and Observer* (May 13, 1945), 9.
7. Adams S, Phillips, board clash on kindergarten data, *Raleigh News and Observer* (March 6, 1975), 1, 20.
8. American Statistical Association, Preparing master's statistics students for success: ASA board approves workgroup recommendations, *Amstat News* (June, 2013), 21.
9. American Statistical Association Undergraduate Guidelines Workgroup, Curriculum guidelines for undergraduate programs in statistical science, American Statistical Association, Alexandria, VA, 2014, `https://www.amstat.org/asa/files/pdfs/EDU-guidelines2014-11-15.pdf`.
10. Anderson RL, Gertrude Mary Cox, *Biographical Memoirs* **59** (1990), 117–132.
11. Cochran WG, Gertrude Mary Cox, 1900–1978, *International Statistical Review* **47** (1979), no. 1, 97–98.
12. Cochran WG, Cox GM, *Experimental designs*, Wiley, New York, 1950.
13. Cox GM, A statistical study of industrial science students of the class of 1926, *Proceedings of the Iowa Academy of Science* **37** (1930), 337–341.
14. Cox GM, A survey of types of experimental designs, *Biometrics* **6** (1950), no. 3, 301–302.
15. Cox GM, Elements of an effective inter-American training program in agricultural statistics, *Estadistica* **11** (1953), 120–128.
16. Cox GM, Statistical frontiers, *Journal of the American Statistical Association* **52** (1957), no. 277, 1–12.
17. Cox GM, Letter to Pat Barber, Gertrude Mary Cox Papers, MC 00117, Special Collections Research Center, North Carolina State University Libraries, Raleigh, NC, 1959.
18. Cox GM, Homeyer PG, Professional and personal glimpses of George W Snedecor, *Biometrics* (1975), 265–301.
19. Cumming D, Gertrude Cox: Statistically-speaking, she's one of a kind, *Raleigh News and Observer* (May 4, 1975), III–3.
20. Grier DA, *When computers were human*, Princeton University Press, Princeton, NJ, 2005.
21. Hall NS, Ronald Fisher and Gertrude Cox: Two statistical pioneers sometimes cooperate and sometimes collide, *The American Statistician* **64** (2010), no. 3, 212–220.
22. Hamblen JW, Statistical programs for the IBM 650—part I, *Communications of the ACM* **2** (1959), no. 8, 13–18.
23. Hotelling H, The teaching of statistics, *The Annals of Mathematical Statistics* **11** (1940), no. 4, 457–470.
24. Lohr SL, Statistical frontiers in survey sampling, *American Statistician* **58** (2004), no. 2, 145–149. MR 2061199
25. Lush JL, Early statistics at Iowa State University, *Statistical Papers in Honor of George W. Snedecor* (T A Bancroft, ed.), The Iowa State University Press, Ames, IA, (1972), pp. 211–226.
26. National Academies of Sciences, Engineering, and Medicine, *Refining the concept of scientific inference when working with big data: Proceedings of a workshop*, National Academies Press, Washington, DC, 2017.
27. Nourse ES, Greenberg BG, Cox GM, Mason DD, Grizzle JE, Johnson NJ, Jones LV, Monroe J, Simons Jr. JD, Statistical training and research: the University of North Carolina system, *International Statistical Review* **46** (1978), no. 2, 171–207.
28. Reeves BC, Scott LJ, Taylor J, Harding SP, Peto T, Muldrew A, Hogg RE, Wordsworth S, Mills N, O'Reilly D, Rogers CA, Chakravarthy U, Effectiveness of community versus hospital eye service follow-up for patients with neovascular age-related macular degeneration with quiescent disease (ECHoES): a virtual non-inferiority trial, *BMJ Open* **6** (2016), no. 7, e010685.
29. Rossiter MW, *Women Scientists in America: Before Affirmative Action, 1940–1972*, Johns Hopkins University Press, Baltimore, MD, 1995.
30. Shetterly ML, *Hidden Figures: The American dream and the untold story of the black women mathematicians who helped win the space race*, William Morrow, New York, 2016.
31. Sylvester DJ, Lohr SL, The security of our secrets: A history of privacy and confidentiality in law and statistical practice, *Denver University Law Review* **83** (2005), 147–209.
32. US Department of Health, Education, and Welfare, *Records, computers and the rights of citizens*, Department of Health, Education, and Welfare, Washington, DC, 1973, `https://aspe.hhs.gov/report/records-computers-and-rights-citizens`.
33. Wood T, Porter E, The elusive backfire effect: Mass attitudes' steadfast factual adherence, *Political Behavior* (2018), 1–29.

ACKNOWLEDGMENT. The author is grateful to the associate editor and reviewers for their helpful comments.



Sharon L. Lohr

Credits

Figure 1 is reproduced with permission from University Archives Photographs, People (ua023_024-001-bx0004-019-009), Special Collections Research Center, North Carolina State University Libraries, Raleigh, North Carolina.

Author photo is courtesy of the author.

The Life and Pioneering Contributions of an African American Centenarian: Mathematician Katherine G. Johnson

Katherine G. Johnson, in her late nineties.

Johnny L. Houston

For permission to reprint this article, please contact: `reprint-permission@ams.org`.

DOI: http://dx.doi.org/10.1090/noti1809

Johnny L. Houston is Executive Secretary Emeritus (1975–2000) of the National Association of Mathematicians. His email address is `jlhouston602@gmail.com`.

Communicated by Notices *Associate Editor Robin Wilson.*

I first met mathematician Katherine G. Johnson when I presented her with the National Association of Mathematicians (NAM) Distinguished Service Award at NAM's Regional Conference in Norfolk, Virgina in 1996. The award celebrated her more than 50 years as a productive mathematician, most of these years having been spent with the National Aeronautics and Space Administration (NASA).

Given her necessary security clearance, NAM did not probe Johnson about the nature of her work. For an African American mathematician to have worked at NASA from the 1950s well into the 1980s was itself historic.

The full and extraordinary story of her life and pioneering contributions was revealed to the world in the *New York Times* bestseller *Hidden Figures*, published in 2016 and written by Margot Lee Shetterly. However, Johnson's contributions became best known after the Oscar®-nominated movie *Hidden Figures* was released in December 2016. I was highly impressed with Johnson from what I learned from both the book and the movie. She was a pioneer extraordinaire and a brilliant mathematician. Her work impacted the success of NASA's early space flight missions. I was extremely delighted that in her lifetime she has received the awards, honors, and recognitions that many pioneers never live long enough to witness and enjoy. I found her life story itself to be a fascinating one to know.

Katherine Coleman was born on August 26, 1918 in White Sulphur Springs (Greenbrier County), West Virginia as the fourth and youngest child of Joshua and Joylette Coleman. She had a brother Horace (b. 1912), a sister Margaret (b. 1913), and a brother Charles (b. 1915). Her mother was a schoolteacher and her father was a lumberman, farmer, and handyman who worked at the Greenbrier Hotel.

At an early age (her third birthday or younger) she began to speak very articulately and was very curious about knowing details of everyday things that she observed. Her father, with only a sixth-grade education, had an incredible ability for doing math problems. For Katherine, he was the smartest person she knew. She started to count everything she saw and attempted to emulate her father in solving math problems. For Katherine, counting things and constantly learning new information about things became her favorite daily activity. Her mother being a nurturing teacher and her father being a math whiz kept her motivated to learn. Katherine officially began attending elementary school at the age of five.

However, because of what she had learned prior to that age, she was placed into the second grade during her first year of school. When she was eight years old she should have entered the fifth grade but, being such an advanced student, she was placed in the sixth grade of a newly opened school for Blacks. With her advanced placement, she was now a grade ahead of her brother Charles who was three years older than she was. At age ten, Katherine was ready to enter high school. She was viewed by many as a child prodigy. Her father instilled in her that she was as good as anyone and could achieve whatever she desired, but she was never to think that she was better than others.

White Sulphur Springs had no high school for Black children. Because the Coleman family valued education highly and was determined that their children should have a quality education through high school and college, her parents rented a house in Institute, Kanawha County, West Virginia where their children attended high school and college. Thus, every autumn for eight years, Katherine's mother moved with her children to the rented home. In the summer they would return some 125 miles back to White Sulphur Springs where her father lived in their *home house* and worked, primarily as a farmer and at a hotel, earning about $100 per month. All four Coleman children completed high school and college under this arrangement of living in two different places during the year.

Katherine entered West Virginia State College High School before her teens and graduated at the age of 14. In high school, she excelled in mathematics, science, and English. In high school she also developed some affinity for astronomy. This was where she met another person who greatly influenced her love for math: Angie Turner King, who taught her geometry in high school. King later taught her math in college and continued to encourage her.

Katherine entered West Virginia State College (WVSC), a Historically Black College (HBCU), in her early teens. As a student at WVSC, she took every math course offered by the college. Several professors mentored her math studies, including chemist and mathematician Angie Turner King, who had also taught her in high school. Katherine said that King was "...a wonderful teacher—bright, caring, and very rigorous." James Carmichael Evans, who had BS and MS degrees from the Massachusetts Institute of Technology, also nurtured Katherine in her study of math. He was a very talented and encouraging teacher who insisted that she must major in mathematics, even though he knew of her strong interests and mentoring in French and English by others. And there was W. W. Schieffelin Claytor, the third African American to receive a PhD degree in math who took Katherine "under his wing." He was a brilliant teacher and researcher. Claytor not only taught her many of her math classes, but he also added new math courses to the curriculum just for Katherine. She recalled that Claytor told her, "You would make a good research mathematician" (after her sophomore year), and he continued, "I am going to prepare you for that career." According to a videotaped interview with Katherine, one of the courses Claytor created for her was analytic geometry, which was invaluable to her in her work at NASA. She was very fortunate to have had Claytor as a teacher. He only taught at West Virginia State College from 1934 to 1937. Katherine graduated from WVSC summa cum laude at the age of 18 in 1937 with degrees in mathematics and French; she had joined Alpha Kappa Alpha sorority while a student.

After graduation from college, she took a teaching job at a Black public school in Marion, Virginia. She was offered the job in Marion because she could teach math, teach French, and play the piano. In 1939, Katherine married James Francis Goble, who was called "Jimmy" by his friends. He worked as a high school chemistry teacher in Marion. This marriage produced three daughters: Constance, Joylette, and Katherine. All three became mathematicians and teachers.

In 1940 (before having children), Katherine enrolled in a graduate math program. She entered the graduate program at West Virginia University in Morgantown, West Virginia, the flagship university for the state of West Virginia that had been reserved for White students only. She was the first African American woman to attend the university's graduate school. This was facilitated with the courage of and assistance from WVSC's president, Dr. John W. Davis. He selected her as one of three African American students (she was the only female) to integrate the graduate school after the United States Supreme Court ruling Missouri *ex rel.* Gaines v. Canada (1938). The court ruled that states that provided public higher education for White students also had to provide it for Black students, to be satisfied either by establishing Black colleges and universities or by admitting Black students to previously White-only universities. Katherine spent a term at the University but left the program after she became pregnant. She chose to give priority, at that time, to raising a family. Jimmy and her parents supported her decision.

Katherine returned to teaching when her three daughters grew older. She taught in Morgantown and Bluefield, West Virginia. However, it was not until 1952 when a relative told her about open positions in mathematics at the all-Black West Area Computing Section at the National Advisory Committee for Aeronautics (NACA), Langley Laboratory, Hampton, Virginia that she desired a different use of her mathematical talent. The program was headed by Dorothy Vaughan, whom she had met some years earlier in West Virginia. Katherine and her husband, Jimmy, decided to move the family to Newport News, Virginia to pursue this opportunity. The NACA had stopped hiring in 1952 when they arrived and she worked as a substitute teacher for a year. Katherine was hired by Langley the next year and began work there in the summer of 1953. Just two weeks into Katherine's tenure in the office, Dorothy Vaughan assigned her to a project in the Maneuver Loads Branch of the Flight Research Division. Katherine's temporary position with the previously all White research team soon became permanent. She spent the next four years analyzing data from flight tests and worked on the investigation of a plane crash caused by wake turbulence. As she was completing this work, her husband Jimmy died from a serious medical challenge in December 1956.

Katherine sang in the choir at Carver Memorial Presbyterian Church in Newport News, Virginia for 50 years. The minister there introduced James A. Johnson to her. He had been commissioned in 1951 as a Second Lieutenant in the United States Army and was a veteran of the Korean War. In 1959, the two married. Katherine had no additional children with her second husband.

Both in the West Area Computing Section and in the Flight Research Division, Katherine worked as a "human computer," doing the complex math calculations for airplanes and space flights. NACA disbanded the "Colored

Katherine G. Johnson working at NASA in the 1960s as a "human computer," physicist, and aerospace technologist in an all-White research division of engineers.

Human Computers Group" in 1958 when it was superseded by NASA, which adopted digital computers. In the Research Flight Division, where she was the only Black, all the Whites were hired as engineers and Katherine was considered a "human computer," a mathematician, a physicist, and an aerospace technologist. During the NACA era, Katherine had to leave the Research Flight Division and go back to the West Area Colored Section to use the restroom, eat, or retrieve something out of her locker. Her questions about her daily inconveniences had a great impact on persuading NASA to eradicate its segregated facilities in the early 1960s. At NASA, she fulfilled Claytor's prophecy and vision. She became a world-class research mathematician on the stage of the largest grand challenge problem of the time: successfully conquering the frontier of space flights to other celestial bodies in space. Katherine G. Johnson made many pioneering contributions on this grand challenge stage. For the sake of brevity of this document, only 15 will be listed.

Katherine G. Johnson's primary contributions at NASA were in computational science and research. In 2017 NASA opened and named in Johnson's honor the above state-of-the-art 40,000 square-foot Computational Research Facility at NASA Langley in Hampton, Virginia.

Fifteen of Katherine G. Johnson's Major Pioneering Contributions to Space Flight History:

A. Katherine Johnson was the first African American and the first woman to work in NASA's Research Flight Division.

B. She was the first African American and the first woman to attend NASA's Research Test Flight Briefings where the fundamental problems of a space flight mission were presented, discussed, and analyzed; she specifically requested to be able attend, and they honored her requests.

C. She was the first African American and first woman to have her name placed on a Scientific Report at NASA; however, she actually did major work on many earlier reports for which she received no written credit or recognition in the report itself. The first report with Katherine's name on it was major for NASA [8]. It contained the theory necessary for launching, tracking, and returning space vehicles and was used for the famous space flight by Alan Shepard in May 1961 and the flight of John Glenn in February 1962.

D. Currently, there are more than twenty five scientific reports in the NASA archive in space flight history that Katherine authored or co-authored, the largest number by any African American or woman.

E. From 1958 until her retirement in 1986, Johnson worked as an aerospace technologist in the Spacecraft Controls Branch where all final decisions were made for space travel; she served as NASA's premier research mathematician at the time.

F. She calculated the trajectory for the May 5, 1961 flight of Alan Shepard, the first American to travel in space.

G. She also calculated the launch window for Shepard's 1961 Mercury mission.

H. She plotted backup navigation charts for astronauts in case of electronic failures.

I. When NASA used electronic computers for the first time to calculate John Glenn's orbit around the Earth, NASA's officials called on Johnson to verify the computer's numbers. Glenn specifically asked for Johnson's verifications, and he refused to fly unless she verified the calculations. These were very difficult calculations; they had to account for the gravitational pulls of celestial bodies.

J. As NASA began relying heavily on digital computers, they used Johnson's calculations to help them check the accuracy of the computers; her validations caused NASA to establish confidence in the new digital computer technology.

K. In 1961, NASA used Johnson's calculations of trajectories to help to ensure that Alan Shepard's Freedom 7 Mercury capsule would be found quickly after landing.

L. Johnson also helped to calculate the trajectory for the 1969 Apollo 11 flight to the Moon.

M. In 1970, Johnson worked on the Apollo 13 moon mission; her work on backup procedures and charts helped set a safe path for the crew's return to Earth.

N. In case of malfunctioning, Johnson had helped to create a one-star observation system that would allow astronauts to determine their location with accuracy.

O. Later in her career, Johnson worked on the Space Shuttle Program, the Earth Resources Satellite, and on plans for a mission to Mars.

In recognition of her life and contributions as a role model, a scholar, an educator, and her pioneering career as a research mathematician with NASA in space travel, Johnson has received many awards, honors, and recognitions. For the sake of brevity, only 20 will be listed.

Having recently celebrated her centennial birthday, Katherine Coleman Goble Johnson has lived to receive many awards, honors, and recognitions for her pioneering work. One such award was the Presidential Medal of Freedom from President Barack Obama in 2015.

Twenty of Katherine G. Johnson's Awards, Honors, and Recognitions:

A. 2019 (January 18) the National Association of Mathematicians, NAM's Centenarian Award

B. 2018 (August 25) West Virginia University, Morgantown, unveiled a life-size bronze statue of Katherine Johnson on campus and established a STEM scholarship in her name

C. 2018 (May 12) College of William and Mary awarded her an Honorary Doctorate Degree

D. 2017 (September 22) The Katherine G. Johnson Computational Research Facility at NASA Langley in Hampton, Virginia opened and was named in her honor (40,000 sq. feet)

E. 2017 received Daughters of the American Revolution Medal of Honor

F. 2016 Oscar®-nominated movie *Hidden Figures* profiled her life as a "colored human computer" and a research mathematician at NASA

G. 2016 received Presidential Honorary Doctorate of Humane Letters from West Virginia University, Morgantown

H. 2016 *New York Times* bestseller *Hidden Figures*, by Margot Lee Shetterly, profiled her life as a scholar, an educator, a "colored human computer," and a research mathematician at NASA

I. 2016 received the Space Flight Industry Silver Snoopy Award from Leland Melvin

J. 2016 received the Astronomical Society of the Pacific's Arthur B. C. Walker II Award

K. 2016 listed as one of the 100 most influential women worldwide by the BBC

L. 2015 received National Center for Women and Information Technology's Pioneer in Tech Award

M. 2015 received the Presidential Medal of Freedom from then president Barack Obama

N. 2014 received the De Pinza Honor from National Women History's Museum

O. 2012 selected as a Science History Maker (now archived in the Library of Congress)

P. 2010 received an Honorary Doctor of Science from Old Dominion University, Norfolk, Virginia

Q. 2006 received an Honorary Doctor of Science from Capitol University, Laurel, Maryland

R. 1999 selected as West Virginia State College Outstanding Alumnus of the Year

S. 1998 received an Honorary Doctor of Law from SUNY, Farmingdale, New York

T. 1996 received the National Association of Mathematicians Distinguished Service Award

U. 1971, 1980, 1984, 1985, and 1986 received NASA Langley Research Center Special Achievement Award

In Her Own Words: Quotes from Katherine G. Johnson

A. I like to learn. That's an art and a science.

B. Let me do it. You tell me when you want it and where you want it to land, and I'll do it backwards and tell you when to take off.

C. Girls are capable of doing everything men are capable of doing. Sometimes they have more imagination than men.

D. We will always have STEM with us. Some things will drop out of the public eye and will go away, but there will always be science, engineering, and technology. And there will always be mathematics.

E. I don't have a feeling of inferiority. I never had one. I'm as good as anybody, but not better.

F. Like what you do, and then you will do your best.

On August 26, 2018, Katherine Coleman Goble Johnson completed her 100th trip around the Sun, becoming a highly distinguished centenarian African American mathematician. Katherine G. Johnson lives in Hampton, Virginia. She continues to encourage her grandchildren and students to pursue careers in science, technology, engineering, and mathematics (STEM).

References

[1] Chelsea Gohd, "Katherine Johnson, trailblazing NASA mathematician, celebrates 100 trips around the Sun" Space.com. Available at: https://www.space.com/41638-katherine-johnson-celebrates-100th-birthday.html.

[2] "Katherine Johnson, NASA pioneer and 'computer.'" *What Matters*. Available at: https://www.youtube.com/watch?v=r8gJqKyIGhE.

[3] "Katherine Johnson biography." NASA. Available at: https://www.nasa.gov/content/katherine-johnson-biography.

[4] "Katherine Coleman Goble Johnson." MacTutor History of Mathematics archive. Available at: www.history.mcs.st-and.ac.uk/Biographies/Johnson_Katherine.html.

[5] "Katherine Johnson." Wikipedia. Available at: https://en.wikipedia.org/wiki/Katherine_Johnson.

[6] "Katherine Johnson: American mathematician." *Encyclopedia Britannica*. Available at: https://www.britannica.com/biography/Katherine-Johnson-mathematician.

[7] Margot Lee Shetterly, *Hidden Figures*, William Morrow, New York, 2016.

[8] T H Skopinski and Katherine G Johnson, Determination of Azimuth Angle at Burnout for Placing a Satellite over a Selected Earth Position, NASA Scientific Report, Available at: https://ntrs.nasa.gov/search.jsp?R=19980227091.

Johnny L. Houston

Credits

Opening photo is from NASA, public domain.

Photos of Johnson working at NASA in the 1960s left, by NASA_LANGLEY [CC BY-NC-SA 2.0 (https://creativecommons.org/licenses/by-nc-sa/2.0/)], from flickr.com; right, is in the public domain.

Photo of NASA facility named for Johnson is by Johnny L. Houston.

Photo of Johnson at 2015 Presidential Medal of Freedom ceremony is in the public domain.

Author photo is courtesy of Johnny L. Houston.

The Mathematics of Grace Murray Hopper

Asher Auel

Grace Murray Hopper (1906–1992) is well known as a pioneering computer scientist and decorated Naval officer.

Asher Auel is Gibbs Assistant Professor of Mathematics at Yale University. His email address is asher.auel@yale.edu.

Communicated by Notices *Associate Editor Angela Gibney.*

For permission to reprint this article, please contact: reprint-permission@ams.org.

DOI: https://doi.org/10.1090/noti1810

Her achievements read as a list of firsts: she was an expert at programming Harvard's Mark I, the first large-scale electromechanical computing machine; she was part of the team who developed the UNIVAC I, the first commercial computer produced in the United States, for which she wrote the first compiler; she created the first English-based data processing language FLOW-MATIC, a principal precursor for COBOL, one of the most important programming languages for business applications; and when she

Figure 1. Grace Hopper at the blackboard with students, 1957.

Figure 2. Grace Hopper teaching a COBOL class, 1961.

Figure 3. Grace Hopper with programmers at the console of UNIVAC I, 1957.

retired from the Navy as a Rear Admiral, at 79 years old, she was the oldest active-duty officer in the entire armed forces. She has been widely lauded for these accomplishments. Named in her honor are: a Naval guided missile destroyer warship; a super computer at the National Energy Research Scientific Computing Center; several buildings and a bridge on Naval bases; a park in Arlington, Virginia; a major yearly convention for women in computer science and technology; several prizes, including an early career award from the Association for Computing Machinery; and a recently renamed residential college at Yale University, among others. Her inspiring story has been the subject of many books and several upcoming film projects.

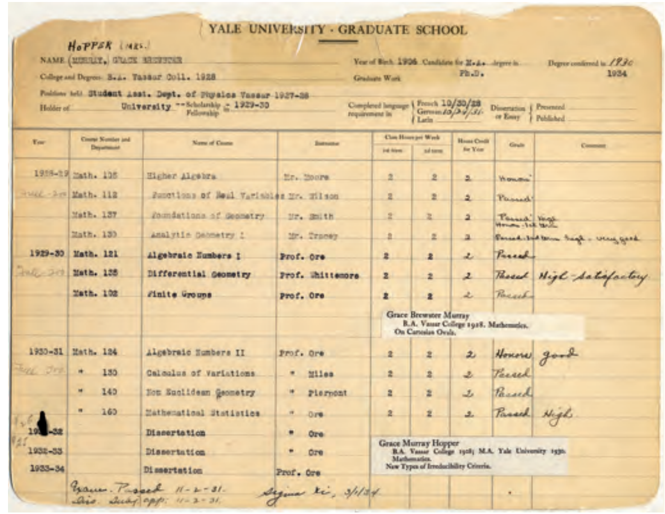
However, what is often overlooked in accounts of Hopper's life and work is her mathematical legacy. The results of her 1934 Yale PhD thesis advised by Øystein Ore (which are detailed in the section "Thesis Work") are never mentioned. Incorrect characterizations of her graduate work abound; her PhD is routinely cited as being in "mathematics and physics" or "mathematical physics" or "under computer pioneer Howard Engstrom." Her training in pure mathematics and her identity as a mathematician are often minimized or treated as a kind of incongruous early chapter in the story of the "Queen of Code."

But Grace Hopper was most certainly a mathematician. Asked in an interview [30, p. 7] later in her career what she would consider herself, she immediately replied: "Mathematician." Then adding wryly: "A rather degraded one now, because I deal with actual digits instead of letters and formulas." Her broad and rigorous mathematical education constituted what she called her "basic thinking." She was, once and forever, a mathematician: "I've been called an engineer, a programmer, systems analyst and everything under the sun but I still think my basic training is mathematics." For the first time, using archival material from Yale University's collections, this article will attempt to illuminate Hopper's foundational mathematical training as well as the specific contributions of her thesis research.

Academic Training

As both an undergraduate and a graduate student, Grace Hopper pursued a mathematical education. In 1928, she earned her BA from Vassar College, with her coursework primarily in mathematics, and secondarily split between economics and physics. She then enrolled as a graduate student in the Department of Mathematics at Yale University, receiving her MA in 1930 with a thesis titled *On Cartesian Ovals* and her PhD in 1934 with a dissertation titled *New Types of Irreducibility Criteria*. Hopper took courses in a wide variety of fields, as her graduate transcript reveals (see Figure 5). Her PhD advisor was Norwegian algebraist Øystein Ore, who had recently been recruited to Yale and "breathed new life into an aging department" [32, p. 10].

Figure 4. Grace Murray's senior portrait from the Vassar yearbook, 1928.



YALE UNIVERSITY - GRADUATE SCHOOL

HoPPER (MRS.)
NAME (MURRAY,) GRACE BREWSTER
Year of Birth 1906 Candidate for M.A. degree in Ph.D. Degree conferred in 1930 1934
College and Degree B.A. Vassar Coll. 1928
Graduate Work
Positions held Student Asst. Dept. of Physics Vassar 1927-28
Holder of University --Scholarship Fellowship -- 1929-30
Completed language requirement in French 10/30/28 German 10/24/31 Latin
Dissertation or Essay Presented Published

Year	Course Number and Department	Name of Course	Instructor	Class Hours per Week 1st term	2d term	Hours Credit for Year	Grade	Comment
1928-29	Math. 135	Higher Algebra	Mr. Moore	2	2	2	Honors	
	Math. 112	Functions of Real Variables	Mr. Wilson	2	2	2	Passed	
	Math. 137	Foundations of Geometry	Mr. Smith	2	2	2	Passed	
	Math. 130	Analytic Geometry I	Mr. Tracey	2	2	2		
1929-30	Math. 121	Algebraic Numbers I	Prof. Ore	2	2	2	Passed	
	Math. 125	Differential Geometry	Prof. Whittemore	2	2	2	Passed	High-Satisfactory
	Math. 102	Finite Groups	Prof. Ore	2	2	2	Passed	
1930-31	Math. 124	Algebraic Numbers II	Prof. Ore	2	2	2	Honors	good
	" 130	Calculus of Variations	" Miles	2	2	2	Passed	
	" 140	Non Euclidean Geometry	" Pierpont	2	2	2	Passed	
	" 160	Mathematical Statistics	" Ore	2	2	2	Passed	High
1931-32		Dissertation	" Ore					
1932-33		Dissertation	" Ore					
1933-34		Dissertation	Prof. Ore					

Grace Brewster Murray
B.A. Vassar College 1928. Mathematics.
On Cartesian Ovals.

Grace Murray Hopper
B.A. Vassar College 1928; M.A. Yale University 1930.
Mathematics.
New Types of Irreducibility Criteria.

Exam. Passed 11-2-31.
Diss. Subj. app. 11-2-31.
Sigma Xi, 3/1/34.

Year	Course Name	Instructor
1928–29	Higher Algebra Foundations of Real Variables Foundations of Geometry Analytic Geometry I	Lucius Terrell Moore Wallace Alvin Wilson Percey Franklyn Smith Joshua Irving Tracey
1929–30	Algebraic Numbers I Differential Geometry Finite Groups	Øystein Ore James K. Whittemore Øystein Ore
1930–31	Algebraic Numbers II Calculus of Variations Non Euclidean Geometry Mathematical Statistics	Øystein Ore Egbert J. Miles James P. Pierpont Øystein Ore
1931–32	Dissertation	Øystein Ore
1932–33	Dissertation	Øystein Ore
1933–34	Dissertation	Øystein Ore

Figure 5. Grace Murray Hopper's original Yale Graduate School transcript lists: her courses and the grades she received (on an Honors/High Pass/Pass/Fail scale), the dates of her language exams (which years later became fodder for her stories about the interchangeability of written languages [30, pp. 22–23]), as well as her yearly tuition and the date of her election to Sigma Xi, the scientific research honor society. Transcription includes full names of her professors.

Notable on Hopper's transcript are Ore's courses on Algebraic Numbers, which had never been offered until his arrival the previous year [32, p. 10]. Hopper was awarded numerous prestigious dissertation fellowships during her years at Yale and was one of the first dozen women (going back to 1895) to earn doctoral degrees in mathematics from the university, see [16]. In 1931, while still a graduate student, Hopper started a faculty position at Vassar, eventually being promoted to assistant professor in 1939 and associate professor in 1944. During the 1941–42 academic year, Hopper was granted a half-time leave from Vassar to take courses with Richard Courant at New York University's Center for Research and Graduate Education (later to become the Courant Institute of Mathematical Sciences).

Numerous distinguished mathematicians can be counted as Hopper's mentors. At Vassar, she studied with Henry Seely White (1861–1943), a prominent American geometer who received his PhD under Felix Klein in 1891 and served as President of the AMS (1906–1908), and Gertrude Smith (1874–1965), whom Hopper declared "taught the best calculus anybody ever taught" [30, p. 21]. At Yale, she was influenced by James P. Pierpont (1866–1938) and was a close contemporary of Howard Engstrom (1902–1962), who received his PhD under Ore in 1929, five years before Hopper, and who eventually returned to Yale to take up a faculty position. In 1941, Engstrom joined the Navy and did foundational work in cryptography; later, he became a deputy director of the National Security Agency, see [12].

Engstrom encouraged several mathematicians, including Hopper and the famous group theorist Marshall Hall, Jr., to join Naval intelligence during World War II. Hall, who received his PhD under Ore in 1936, recalls that while Ore was his "nominal" advisor, he received "far more help and direction" from Engstrom, see [18]. In interviews over the years, Hopper repeatedly describes Engstrom as one of her "instructors." Though no course with Engstrom is listed on Hopper's transcript, his mentorship seems to have been as important for Hopper as it was for Hall. What is clear from the historical record is that Hopper did not "receive her PhD under Engstrom" as several authors have claimed (see [16]), perhaps in an effort to link the early histories of two pioneers in the field of computers. When Hopper enlisted in the Navy, she expected to be assigned to the Communications Supplementary Activity (Navy Communications Annex) in Washington, DC, where Engstrom led a top-secret team building cryptographic computing machines. Though she was eventually assigned to work on the Mark I at Harvard, Hopper and Engstrom stayed life-long friends.

Several of Ore's graduate students from the early 1930s tackled similar research problems on generalized irreducibility criteria for polynomials. His two male students, Harold Dorwart (PhD 1931) and Casper Shanok (PhD 1933), produced dissertations very close in subject to Hopper's work. Dorwart published a number of articles from his thesis work: in the *Annals of Mathematics* [7], in the *Duke Mathematical Journal* [8], and a survey in the *American Mathematical Monthly* [9, p. 373] that mentions and cites Hopper. Almost immediately following his graduation, Shanok's thesis [45] was published in the *Duke Mathematical Journal* (though he did not appear to continue in academia). Distressingly, Ore's two female graduate students, Hopper and Miriam Becker (PhD 1934), never published their thesis work at all. Both would, however, go on to long careers (Becker would eventually join the faculty of the City University of New York), even if their earliest work still remains unknown.

As a junior faculty member at Vassar, Hopper was given "all the courses nobody else wanted to teach." But she was such an innovative teacher that classes like technical drawing, trigonometry, calculus, probability, and finite difference method for numerical solutions of differential equations were suddenly popular [30, pp. 16–21]. On top of her demanding teaching schedule of five or six courses, she also audited two courses per year, including basic astronomy, statistical astronomy, geology, philosophy, bacteriology, biology, zoology, plant horticulture, chemistry, physics, economics, and architecture. She also took a course on cryptography sponsored by the Navy [30, p. 27].

Later in life, Hopper would reflect on the "inestimable value" of her broad education as she shaped the new field of computers [30, p. 17]. For example, it was in a chemistry course when she learned the essential concepts of round-off and truncation errors [27, p. 46]. Her years teaching technical drawing courses enabled her to invent a new method for diagramming the relay timing and associated circuitry (see Figure 6) for the Mark I (formally known as the Automatic Sequence Controlled Calculator) control manual [22], see [30, pp. 32–33]. Hopper summed it up quite neatly in a 1986 interview on *The Late Show with David Letterman* [21]. During a discussion of her Mark I days, Letterman asked, "Now, how did you know so much about computers then?" "I didn't," Hopper immediately replied, with some bemusement. "It was the first one."

But arguably, it was studying and teaching mathematics —thinking about symbolic language and how to communicate meaning with symbols—that was most pivotal in Hopper's early work on computers. Her invention of various types of early compilers enabled the translation of mathematical statements or English words into computer code.

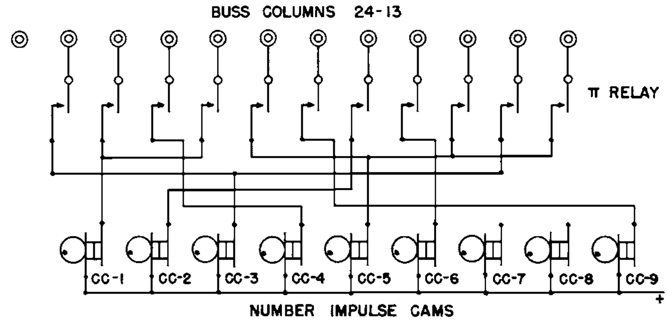


Figure 6. Wiring diagram of a Mark I table relay encoding π [22, p. 91].

> Manipulating symbols was fine for mathematicians but it was no good for data processors who were not symbol manipulators. Very few people are really symbol manipulators. If they are they become professional mathematicians, not data processors. It's much easier for most people to write an English statement than it is to use symbols. So I decided data processors ought to be able to write their programs in English, and the computers would translate them into machine code. [13, p. 3]

One of Hopper's most academically rewarding experiences was taking courses from Richard Courant at New York University in 1941–1942, during her half-time leave funded by a Vassar Faculty Fellowship. Hopper found Courant to be "one of the most delightful people to study with I've ever known in my life." It was, she recalled, "a perfectly gorgeous year. Of course, he scolded me at intervals, just as all of the others did because I kept doing unorthodox things and wanting to tackle unorthodox problems" [30, p. 28]. While there, she studied calculus of variations, differential geometry, and perhaps most fortuitously, she took a government-sponsored defense training course on methods of solutions to partial differential equations involving finite differences taught by Courant, see [30, p. 24]. Hopper later learned that her involvement in this course was in her Navy file and was one of the determining factors in her initial assignment: to program Harvard's Mark I, implementing calculations for the war effort including some for John von Neumann's work on the Manhattan Project.

The attack on Pearl Harbor, which took place during her year studying with Courant, forever changed the direction of Hopper's life. Her great grandfather had been in the Navy, and by the summer of 1942, many of Hopper's family members were joining the armed services: her husband (from whom she was already separated) and brother volunteered for the draft; her female cousins joined through the Women's Army Corps (WAC) and the Navy's Women Accepted for Volunteer Emergency Service (WAVES) program; her mother served on the Ration Board; and her retired father went back to work and served on the local Draft

Figure 7. Grace Hopper standing behind a car parked near Cruft Lab, Harvard University, ca. 1945–1947.

Board, see [47, p. 20]. Hopper was eager to enlist in the Navy, but was rejected when she failed to meet the minimum weight requirement for her height and was considered too old for enlistment. In the meantime, she taught an accelerated summer calculus course at Barnard College for women training for war-related posts. But her profession was also an impediment.

> Mathematicians were [in] an essential industry and you could not leave your job to go in the services without permission [from both the Navy and one's employer]. You couldn't even transfer jobs without permission... And I was beginning to feel pretty isolated sitting up there, the comfortable college professor—all I was doing was more teaching, and I wanted very badly to get in and so I finally gave Vassar an ultimatum that if they wouldn't release me I would stay out of work for six months because I was going into the Navy, period. [30, p. 25]

Eventually, she obtained a waiver for the weight requirement and a leave of absence from Vassar, and trained at the Naval Reserve Midshipmen's School at Smith College in Northampton, Massachusetts in the spring of 1944. After graduating first in her class, she was commissioned lieutenant junior grade.

On July 2, 1944, Hopper reported for duty at the Bureau of Ships Computation Project at Harvard under the command of Howard Aiken, and began work on the Mark I. Aside from programming the Mark I, and its successor, the Mark II, she was assigned the job of compiling notes about the operation of the Mark I into a book [22]. Hopper

Figure 8. Captain Grace Hopper, ca. 1975.

edited the volume and wrote several of its sections, including an introduction containing the first ever scholarly account of the history and development of calculating machines [22, Chapter I]. "Nobody had done this before," Hopper later said. "[The] history of computers had never been put together." It was, to use her words, "really a job" [30, p. 32].

Thesis Work

Grace Hopper's PhD thesis work with Øystein Ore concerned irreducibility criteria for univariate polynomials over the field of rational numbers. Though her work was never published, it was presented to an American Mathematical Society meeting on March 30, 1934 in New York with an abstract appearing in the *Bulletin of the AMS* [20]. The only apparent extant text of her thesis [19] remains in Yale's archives, and a detailed account of her mathematical work has never before appeared in the literature.

In this section, we provide an explanation of Grace Hopper's thesis work, the central theme of which concerns necessary conditions for the irreducibility of univariate polynomials with rational coefficients based on their Newton polygons, see the subsection "Irreducibility via Newton polygons." The connection between the decomposability of polynomials and the slopes of their Newton polygons was initiated at the turn of the twentieth century by Dumas [10], with further refinements by Kürschak [26], Ore [35], and Rella [41]. In her work, Hopper obtains new irreducibility criteria by considering an Archimedean analogue of the Newton polygon, see the subsection "Archimedean Newton polygon." While this Archimedean Newton polygon dates back at least to an 1893 paper of Hadamard [17, §4, p.174], and was later developed further in a 1940 paper of Ostrowski [36, pp. 106, 132] and by Valiron [46, Ch. IX, pp. 193–202], its use for establishing irreducibility criteria seems to be a novel feature of Hopper's work.

Irreducibility of polynomials. A nonconstant polynomial

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

whose coefficients $a_0, \dots, a_n$ are rational numbers is called *irreducible* if there is no way to write $f(x)$ as a product $f(x) = g(x)h(x)$ where $g(x)$ and $h(x)$ are themselves nonconstant polynomials with rational coefficients. The study of irreducible polynomials is one of the foundations of modern field theory and often involves quite a bit of number theory as well.

For example, that $x^2 - 2$ is irreducible is equivalent to the classical fact that $\sqrt{2}$ is irrational. More generally, a quadratic polynomial

$$ax^2 + bx + c$$

is irreducible if and only if its *discriminant* $b^2 - 4ac$, which appears in the quadratic formula, is not a square.

When p is an odd prime number, that the cyclotomic polynomial

$$\Phi_p(x) = x^{p-1} + x^{p-2} + \cdots + x + 1$$

is irreducible was first proved by Gauss in *Disquisitiones Arithmeticae* and is related to the arithmetic of the pth roots of unity $e^{2\pi i k/p}$ and the (non-)constructibility of the regular p-gon with compass and straightedge.

In 1929, Schur [44] proved that for $n \geq 1$, the truncated exponential series

$$1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!}$$

is irreducible with an argument that used a generalization of Bertrand's Postulate, whose original statement—that for any positive integer k there exists a prime number p such that $k < p \leq 2k$—was conjectured by Bertrand and proved by Chebyshev. This result, and its generalizations, implies the irreducibility of various families of orthogonal polynomials, such as those of Laguerre and Hermite type, see [9, §4].

Several standard methods for testing irreducibility are taught in a basic course on field and Galois theory. The most elementary are reduction modulo a prime number and the "rational root test." A more powerful, and yet easy to use, tool is *Eisenstein's criterion*: assuming that $f(x)$ has integer coefficients, if for some prime number p, the coefficients satisfy $p | a_i$ for all $i \neq n$, as well as $p \nmid a_n$ and $p^2 \nmid a_0$, then $f(x)$ is irreducible. In fact, a statement equivalent to Eisenstein's Criterion was first proved by Schönemann [43] in a 1846 paper that Eisenstein even cites in his own paper [11] in 1850, hence the criterion was often called the Schönemann–Eisenstein theorem in literature from the early twentieth century, see [5] for a discussion.

Irreducibility via Newton polygons. In the late nineteenth century and early twentieth century, various generalizations of the Eisenstein criterion, depending on the divisibility properties of the coefficients of $f(x)$, appeared in work of Königsberger, Netto, Bauer, Perron, Ore, and Kahan. Finally, these were all mostly subsumed by an observation of Dumas [10], that such criteria could be rephrased in terms of the irreducibility of the Newton polygon associated to $f(x)$. This history is very well summarized in the historical introduction to Hopper's thesis [19, Chapter I] and in Dorwart's survery article [9].

Given a prime number p, we consider the p-adic valuation v_p on $\mathbb{Q}$. The *Newton polygon* $N_p(f)$ of the polynomial $f(x) = \sum_i a_i x^i \in \mathbb{Q}[x]$ with respect to p is the lower convex hull of the points $(i, v_p(a_i))$ in $\mathbb{R}^2$. We assume that $a_0 \neq 0$. If $a_i = 0$ for some $i \geq 0$ then by definition $v_p(a_i) = +\infty$, hence for the purposes of taking the lower convex hull, we can ignore such zero coefficients. Intuitively, we can imagine a large rubber band surrounding these points in $\mathbb{R}^2$, which each have small nails sticking up from them; as we stretch the rubber band up toward $+\infty$, we obtain the Newton polygon as the lower sequence of line segments formed by the stretched rubber band.

The central insight of Dumas [10, p. 217] is that the Newton polygon $N_p(g \cdot h)$ of the product of polynomials $g(x)$ and $h(x)$ is formed by composing the line segments of the Newton polygons $N_p(g)$ and $N_p(h)$ in order of increasing slope, an operation that we could denote $N_p(g) \circ N_p(h)$ and call the *Dumas sum*. This was generalized in [3] and [6], and by many later authors, including to the more general context of (multivariate) polynomials over valued fields.

If the projections to the x- and y-axes of the line segments of the Newton polygon of $f(x)$ are denoted $l_1, \dots, l_r$ and $k_1, \dots, k_r$, respectively, we denote by $e_i = \gcd(l_i, k_i)$ and write $l_i = e_i \lambda_i$. Then Dumas [10, p. 237] deduces a general irreducibility criterion: $f(x)$ can only have factors

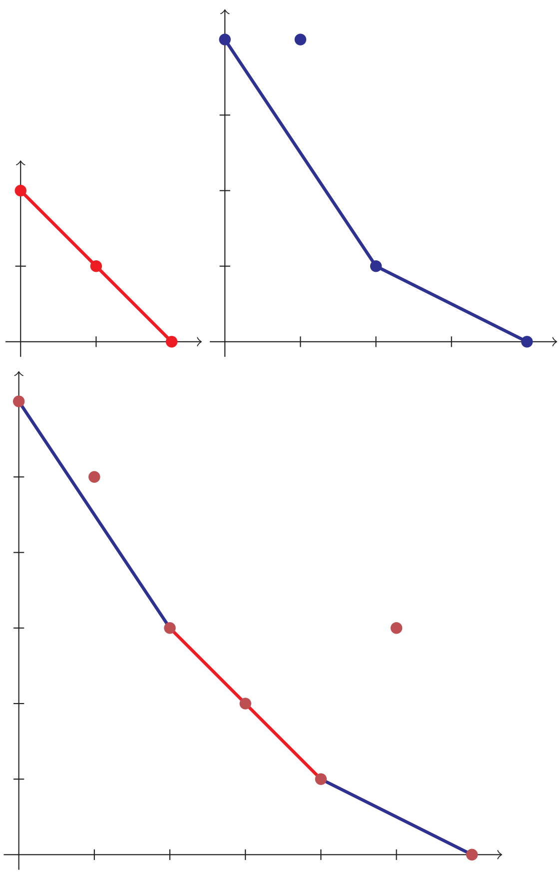


Figure 9. The Newton polygon (with $p = 2$) of $g(x) = x^2 + 6x + 4$ in red, of $h(x) = x^4 + 2x^3 + 10x^2 + 48x + 16$ in blue, and $g(x)h(x) = x^6 + 8x^5 + 26x^4 + 116x^3 + 344x^2 + 288x + 64$, with sides colored appropriately showing the Dumas sum.

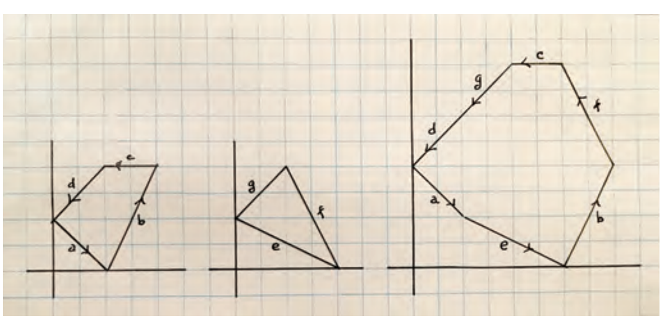


Figure 10. Graphical depiction of the Minkowski sum of two classical Newton polygons of bivariate polynomials. Hand drawn by Grace Hopper [19, p. 24].

of degree m that can be expressed in the form

$$m = \sum_{i=1}^{r} \mu_i \lambda_i$$

where $\mu_i \in \{0, 1, \ldots, e_i\}$ for each $1 \leq i \leq r$.

For example, if $N_p(f)$ consists of a single line segment that does not pass through any lattice point in the plane, then $f(x)$ is irreducible. This immediately gives the Eisenstein criterion. Generalizations and refinements of this idea were developed by Fürtwangler, Kürschak, and Ore, see [19, Chapter I, §4].

The classical Newton polygon associated to a bivariate polynomial $f(x, y)$ over a field, defined as the convex hull of the weight vectors (i, j) in $\mathbb{R}^2$ of all monomials $x^i y^j$ appearing with nonzero coefficients in $f(x, y)$, first appears in a 1676 letter from Newton to Oldenberg [23] and was well known to Newton and his followers throughout the 18th and 19th century, cf. [4, Chapter XXX, §24, Historical Note]. Though it must have been well known, the observation that the classical Newton polygon of a product of polynomials is the Minkowski sum (see Figure 10) of their classical Newton polygons does not seem to be clearly enunciated in the literature until the theses of Shanok [45, §2, p. 103, footnote 3] and Hopper [19, Chapter II, §1].

Archimedean Newton polygon. A completely different type of irreducibility criterion depending on the relative magnitudes of the absolute values of the coefficients was introduced by Perron [40]. (We now assume that $f(x)$ is a monic polynomial with coefficients in $\mathbb{Z}$.) These criteria depend on the following simple observation: if $n - 1$ of the (complex) roots of $f(x)$ have absolute value < 1, then $f(x)$ is irreducible. Indeed, if $f(x)$ has a nonconstant factor (which by Gauss's Lemma can be taken to be monic with integer coefficients), then all of its roots will have absolute value < 1, but their product is the (integral) constant term, a contradiction. The resulting irreducibility criterion is, letting $A = |a_0| + \cdots + |a_{n-1}| + 1$: if the coefficients satisfy $|a_{n-1}| > \frac{1}{2}A$, then $f(x)$ is irreducible. There is a similar criterion if all but a pair of complex conjugate roots have absolute value < 1.

To take into account the relative magnitudes of the coefficients, Hopper [19, Chapter III] considers an *Archimedean Newton polygon* associated to a polynomial $f(x)$ with complex coefficients. Define $N_\infty(f)$ to be the lower convex hull of the set of points $(i, -\log|a_i|)$ in $\mathbb{R}^2$. As before, if $a_i = 0$ for some $0 < i < n$, then $-\log|a_i| = +\infty$, so can be ignored for the purposes of taking the lower convex hull. (In fact, Hopper defines the mirror image of this polygon.) This is a natural generalization of the Newton polygon with respect to a prime p considered above. Indeed, the negative absolute logarithm can be considered as an

Archimedean analogue of a valuation; writing $v_\infty(x) = -\log|x|$, then $|x| = e^{-v_\infty(x)}$ is in analogy with the non-Archimedean p-adic absolute value $|x|_p = p^{-v_p(x)}$.

Later in the twentieth century, the Archimedean Newton polygon was, in various guises, used in a variety of contexts, including: by Khovansky [25] (cf. [42]) in an algebraic reformulation of his study of exponential equations and eventually for combinatorial invariants attached to divisors on algebraic varieties; by Mueller and Schmidt [33], [34] for bounding the number of solutions to Thue equations; and by Passare and his collaborators (see e.g., [38], [37, §2.1], [1]) and Mikhalkin (see e.g., [39]) in the theory of amoebas and in tropical geometry. The genesis of the Archimedean Newton polygon going back to Hadamard, as well as most of these later uses, stems from the fact that its geometry is related to the absolute values of the roots of the polynomial.

Taking a different approach, Hopper [19, Chapter III] studies the Archimedean Newton polygon of a product of polynomials, in analogy with Dumas's result in the non-Archimedean case: how do $N_\infty(g)$ and $N_\infty(h)$ compare with $N_\infty(g \cdot h)$? Hopper remarks that if the analogue of Dumas's product result held for N_∞, then irreducibility criteria such as Perron's, which depend on the relative magnitude of the coefficients, would follow immediately. However, $-\log|x|$ is not a valuation as there is an error term in relating $-\log|x+y|$ with $\min(-\log|x|, -\log|y|)$, hence such an exact product formula is not expected. However, Hopper goes on to prove bounds on how far apart $N_\infty(g \cdot h)$ can be from $N_\infty(g) \circ N_\infty(h)$. To state these bounds, if $f(x) \in \mathbb{C}[x]$ is a polynomial of degree $n \geq 1$, we consider $N_\infty(f)$ as a piecewise-linear function of t on the real interval $[0, n]$.

Theorem 1 (Hopper [19, Chapter III, §3–5, pp. 33–38]). *Let $g(x), h(x) \in \mathbb{C}[x]$ be monic polynomials and $n = \deg(g) + \deg(h)$. Then*

$$-\log(1 + \tfrac{n}{2}) \leq (N_\infty(g \cdot h)(t) - (N_\infty(g) \circ N_\infty(h))(t) \leq \log(3 \cdot 2^{t(n-t)})$$

for all $t \in [0, n]$.

More precisely, Hopper establishes an upper bound, as in Theorem 1, that depends on the *sharpness* of the bends in $N_\infty(g) \circ N_\infty(h)$, defined as the (exponential of the) ratio of slopes of consecutive sides. Near very sharp bends, the two polygons are very close; the careful analysis [19, Chapter III, §5] of bends with small sharpness gives the upper bound. She remarks that the "result can however probably be considerably improved upon" due to certain estimates employed in the proof [19, p. 38].

The Newton–Hopper polygon. In [19, Chapter II, §2], Hopper introduces a new construction of a convex polygon associated to a monic polynomial with integer coefficients that takes into account both the divisibility (with respect to a fixed prime p) and the magnitudes of the coefficients. We call this the *Newton–Hopper* polygon $NH_p(f)$ associated to $f(x) = \sum_i a_i x^i \in \mathbb{Z}[x]$. It is defined by writing

$$f(x) = \sum_i \sum_j r_{ij} p^j x^i$$

where $r_{ij} \neq 0$ and satisfy $-p < r_{ij} < p$, and then taking the convex hull of the points (i, j) in $\mathbb{R}^2$. This construction yields a convex polygon whose "lower half" is $N_p(f)$ and whose "upper half" is the upper convex hull of the points $(i, \lfloor \log_p |a_i| \rfloor)$, so that the upper half is approximately $-N_\infty(f)$. The analogous bounds in Theorem 1 hold for the upper half of the Newton–Hopper polygon of a product.

Hopper's strategy [19, Chapter IV] is then to start with a polynomial $f(x) \in \mathbb{Z}[x]$, plot $NH_p(f)$ (in black ink), and then plot (in red and blue ink) the limits of the upper and lower bounds in Theorem 1 away from $NH_p(f)$. Finally, if one can verify that each possible polygon within the region bounded between the (red and blue) limits cannot be decomposed as a Dumas sum of Newton–Hopper polygons (where we formally apply Dumas composition to the upper half and lower half separately) of lower degrees, then $f(x)$ must be irreducible. This observation provides new irreducibility criteria that simultaneously generalize those depending on the divisibility and the magnitudes of the coefficients.

Hopper then proceeds with a careful analysis of various general situations in which this occurs, and then produces families of sparse polynomials that satisfy these criteria. Some of her families in [19, Chapter IV, §5] cannot be proven to be irreducible solely using either divisibility properties or relative magnitude properties of the coefficients on their own. For example, the polynomial

$$f(x) = x^7 \pm (p^{11} + p)x^5 \pm p^4,$$

for any prime $p > 3 \cdot 2^{49/4} > 14,612$ (e.g., $p = 14,621$ is the first such prime), is irreducible. Similarly, the polynomial

$$f(x) = x^9 \pm (p^6 + p)x^3 \pm (p^9 + p^3)x^2 \pm p^3, \quad (1)$$

for any $p > 3 \cdot 2^{81/4} > 3,740,922$ (e.g., $p = 3,740,923$ is the first such prime) is irreducible, see Figure 11. For all primes p below these bounds, a computer algebra system can verify the irreducibility of the above polynomials. Also, the following infinite family of polynomials

$$f(x) = x^n \pm kx^2 \pm lp^{2\nu+1},$$

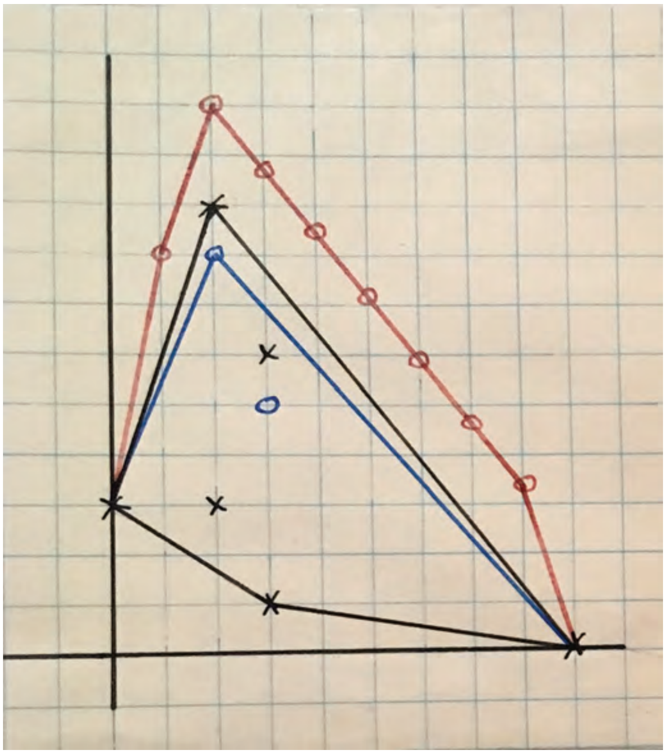


Figure 11. The Newton–Hopper polygon of the polynomial (1), in black, with the upper and lower bounds in Theorem 1 in red and blue, respectively. Hand drawn by Grace Hopper [19, p. 55].

where $n \geq 3$, $\nu \geq 3$, $0 < k < p^{2(\nu-2)}$, $p \nmid k$, $0 < l < p$, and $p > 3 \cdot 2^{n^2/4}$, are all irreducible. Similarly, the following infinite family of polynomials

$$f(x) = x^n \pm kpx \pm mp^\nu,$$

where $n \geq 2$, $\nu \geq 4$, $0 < k < p^{2(\nu-2)}$, $0 < m < p^{\nu-3}$, $p \nmid km$, and $p > 3 \cdot 2^{n^2/4}$, are all irreducible.

ACKNOWLEDGMENTS. The author would like to thank retired Yale Mathematics librarian Paul Lukasiewicz for introducing him to the story of Grace Hopper; head of Grace Hopper College, Julia Adams, for her support and stimulating conversations; Grace Hopper's nephew and great-niece, Roger and Deborah Murray, for their encouragement of this article; Yale University archivist, Michael Lotstein, for his tireless assistance in locating Hopper-related archival material; Vassar College associate registrar, Kathleen Giblin, for providing and deciphering Hopper's college transcript; Sam Payne for productive mathematical discussions; and Jennifer Raab for her generous help during the preparation of this article. Finally, the author is grateful to Judy Green for answering many questions about Hopper's history, and for her pioneering study with Jeanne LaDuke [13], [14], which is an indispensable resource on Hopper's life and work.

References

[1] Avendaño M, Kogan R, Nisse M, Rojas JM, Metric estimates and membership complexity for Archimedean amoebae and tropical hypersurfaces, *Journal of Complexity* **46** (2018), 45–65. MR3774347

[2] Beyer K, *Grace Hopper and the Invention of the Information Age*, MIT Press, Cambridge, MA, 2009.

[3] Blumberg H, On the factorization of expressions of various types, *Trans. Amer. Math. Soc.* **17** (1916), no. 4, 517–544. MR1501056

[4] Chrystal G, *Algebra, Part II*, 2nd ed., Adam & Charles Black, London, 1900. MR0121327

[5] Cox DA, Why Eisenstein proved the Eisenstein criterion and why Schönemann discovered it first, *American Mathematical Monthly* **118** (2011), no. 1, 3–31. MR2795943

[6] Dines LL, A theorem on the factorization of polynomials of certain types, *Bulletin of the AMS* **29** (1923), 440.

[7] Dorwart HL, Ore Ø, Criteria for the irreducibility of polynomials, *Ann. of Math.* (2) **34** (1933), no. 1, 81–94; erratum *Ann. of Math.* (2) **35** (1934), no. 1, 195. MR1503098

[8] Dorwart HL, Concerning certain reducible polynomials, *Duke Math. J.* **1** (1935), no. 1, 70–73. MR1545865

[9] Dorwart HL, Irreducibility of polynomials, *Amer. Math. Monthly* **42** (1935), no. 6, 369–381. MR1523399

[10] Dumas G, Sur quelques cas d'irréductibilité des polynômes coefficients rationnels, *Journal de Mathématiques Pures et Appliquées* (6) **2** (1906), 191–258.

[11] Eisenstein G, Über die Irreductibilität und einige andere Eigenschaften der Gleichung, von welcher die Theilung der ganzen Lemniscate abhängt, *Journal für die reine und angewandte Mathematik* **39** (1850), 160–179. MR1578663

[12] Engstrom HT, Scientist, Was 59: One of the Developers of Univac Computer Dies, *New York Times*, 10 March 1962, p. 21.

[13] Gilbert L, Moore G, Particular Passions: Grace Murray Hopper, chapter in *Women of Wisdom: Talks With Women Who Shaped Our Times*, Lynn Gilbert Inc., 1981.

[14] Green J, LaDuke J, *Pioneering Women in American Mathematics: The Pre-1940 PhD's*, History of Mathematics, vol. 34, American Mathematical Society, Providence, RI, 2008. MR2464022

[15] Green J, LaDuke J, *Supplementary Material for Pioneering Women in American Mathematics: The Pre-1940 PhD's*, available at `http://www.ams.org/publications/authors/books/postpub/hmath-34`. MR2919139

[16] Green J, LaDuke J, Letter to the Editor, *Isis* **102** (2011), no. 1, pp. 136–137. MR1983738

[17] Hadamard J, Étude sur les propriétés des fonctions entières et en particulier d'une fonction considérée par Riemann, *Journal de Mathématiques Pures et Appliquées* (4) **9** (1893), 171–216. MR0220564

[18] Hall M Jr, Mathematical Biography: Marshall Hall Jr., in *A Century of Mathematics in America, Part I, History of Mathematics*, vol. 1, Peter Duran, Richard Askey, Uta

C. Merzbach, eds., American Mathematical Society, Providence, RI, 1988, pp. 367–374. MR1563496
[19] Hopper GM, *New types of irreducibility criteria*, PhD dissertation, Yale University, May 1934, available at `http://math.yale.edu/grace-murray-hopper`. MR2936986
[20] Hopper GM, Ore Ø, New types of irreducibility criteria, *Bull. Amer. Math. Soc.* **40** (1934), abstract no. 126, 216. MR2936986
[21] *Late Night with David Letterman*, season 5, episode 771, October 2, 1986, National Broadcasting Company, New York, NY.
[22] *A Manual of Operation for the Automatic Sequence Controlled Calculator*, by the staff of Computation Laboratory, Annals of the Computation Laboratory of Harvard University, vol. 1, Harvard University Press, Cambridge, MA, 1946. MR0020856
[23] Newton I, letter to Henry Oldenburg dated 1676 Oct 24, in *The Correspondence of Isaac Newton: Volume 2, 1676–1687*, pp. 126–127, Cambridge University Press, 1960. MR551872
[24] Grace Murray Hopper 1906–1992, *Notices Amer. Math. Soc.* **39** (1992), 320. MR1153173
[25] Hovansky A, Sur les racines complexes des systèmes 'déquations algébriques comportant peu de termes, *C. R. Acad. Sci. Paris Ser. I Math.* **292** (1981), no. 21, 937–940. MR625726
[26] Kürschák J, Irreduzible Formen, *Journal für die Mathematik* **152** (1923), 180–191. MR1581009
[27] Grace Murray Hopper (1906–1992), interview by Beth Luebbert and Henry Tropp, 5 July 1972, Computer Oral History Collection, Archives Center, National Museum of American History, Smithsonian Institution, transcript available at `http://amhistory.si.edu/archives/AC0196_hopp720507.pdf`
[28] MacLane S, A construction for absolute values in polynomial rings, *Trans. Amer. Math. Soc.* **40** (1936), no. 3, 363–395. MR1501879
[29] Mathews S, Mogensen M, Grace Brewster Murray Hopper, student paper for Appalachian State University course Women and Minorities in Math, taught by Sarah J. Greenwald, Spring 2001, available at `http://www.cs.appstate.edu/~sjg/wmm/student/hopper/hopperp.htm`.
[30] Grace Murray Hopper (1906–1992), interview by Uta C. Merzbach, July 1968, Computer Oral History Collection, Archives Center, National Museum of American History, Smithsonian Institution, transcript available at `http://amhistory.si.edu/archives/AC0196_hopp680700.pdf`.
[31] Voice of America interviews with eight American women of achievement: Grace Hopper, Betty Friedan, Nancy Landon Kassebaum, Mary Calderone, Helen Thomas, Julia Montgomery Walsh, Maya Angelou, Nancy Clark Reynolds, interviews by Chantal Mompoullan, Voice of America, United States Information Agency, Washington, DC, 1985.
[32] Mostow GD, *Science at Yale: Mathematics*, Yale University Press, New Haven, CT, 2001.
[33] Mueller J, Schmidt WM, Thue's equation and a conjecture of Siegel, *Acta Math.* **160** (1988), no. 3–4, 207–247. MR945012
[34] Mueller J, Schmidt WM, On the Newton Polygon, *Monatshefte für Mathematik*, **113** (1992), no. 1, 33–50. MR1149059
[35] Ore O, Zur Theorie der Irreduzibilitätskriterien, *Mathematische Zeitschrift* **18** (1923), 278–288. MR1544631
[36] Ostrowski A, Recherches sur la méthode de Graeffe et les zéros des polynomes et des séries de Laurent, *Acta Math.* **72** (1940), 99–155. MR0001944
[37] Passare MJ, Rojas M, Shapiro B, New multiplier sequences via discriminant amoebae, *Mosc. Math. J.* **11** (2011), no. 3, 547–560. MR2894430
[38] Passare M, Rullgård H, Amoebas, Monge-Ampère measures, and triangulations of the Newton polytope, *Duke Math. J.* **121** (2004), no. 3, 481–507. MR2040284
[39] Mikhalkin G, Enumerative tropical algebraic geometry in $\mathbb{R}^2$, *Journal of the American Mathematical Society* **18** (2005), no. 2, 313–377. MR2137980
[40] Perron O, Neue Kriterien für die Irreduzibilität algebraischer Gleichungen, *Journal für Mathematik* **132** (1907), 288–307. MR1580727
[41] Rella T, Ordnungsbestimmungen in Integritätsbereichen und Newtonische Polygone, *Journal für die Mathematik* **158** (1927), 33–48. MR1581128
[42] Risler JJ, Complexité Géométrie Réelle, *Sém. Bourbaki (1984-85)*, no. 637, Astérisque **133–134** (1986), 89–100.
[43] Schönemann T, Von denjenigen Moduln, welche Potenzen von Primzahlen sind, *Journal für die reine und angewandte Mathematik* **32** (1846), 93–118. MR1578516
[44] Schur I, Einige Sätze über Primzahlen mit Anwendungen auf Irreduzibilitätsfragen I, *Sitzungsberichte Preuss. Akad. Wiss. Phys.-Math. Klasse* **14** (1929), 125–136. Also in *Gesammelte Abhandlungen*, Band III, 140–151. MR0462893
[45] Shanok C, Convex polyhedra and criteria for irreducibility, *Duke Math. J.* **2** (1936), no. 1, 103–111. MR1545909
[46] Valiron G, *Fonctions analytiques*, Presses Universitaires de France, Paris, 1954. MR0061658
[47] Williams K, *Grace Hopper: Admiral of the Cyber Sea*, Naval Institute Press, Annapolis, MD, 2004.

Asher Auel

Credits

Figures 1 and 2, Hagley Museum and Library "Sperry Corporation, UNIVAC Division" collection of photographs and audiovisual materials. Reproduction is courtesy of the Unisys Corporation.

Figures 3 and 8 are courtesy of the Division of Medicine and Science, National Museum of American History, Smithsonian Institution.

Figure 4 is courtesy of Archives and Special Collections, Vassar College Library.

Figure 5 is courtesy of Graduate School of Arts and Sciences, Yale University, Student Records (RU 262), Manuscripts and Archives, Yale University Library. Reproduction in consultation with family representatives Roger and Deborah Murray.

Figure 6 is courtesy of Harvard University Press. Reproduction authorized and permitted upon publication by the President and Fellows of Harvard College.

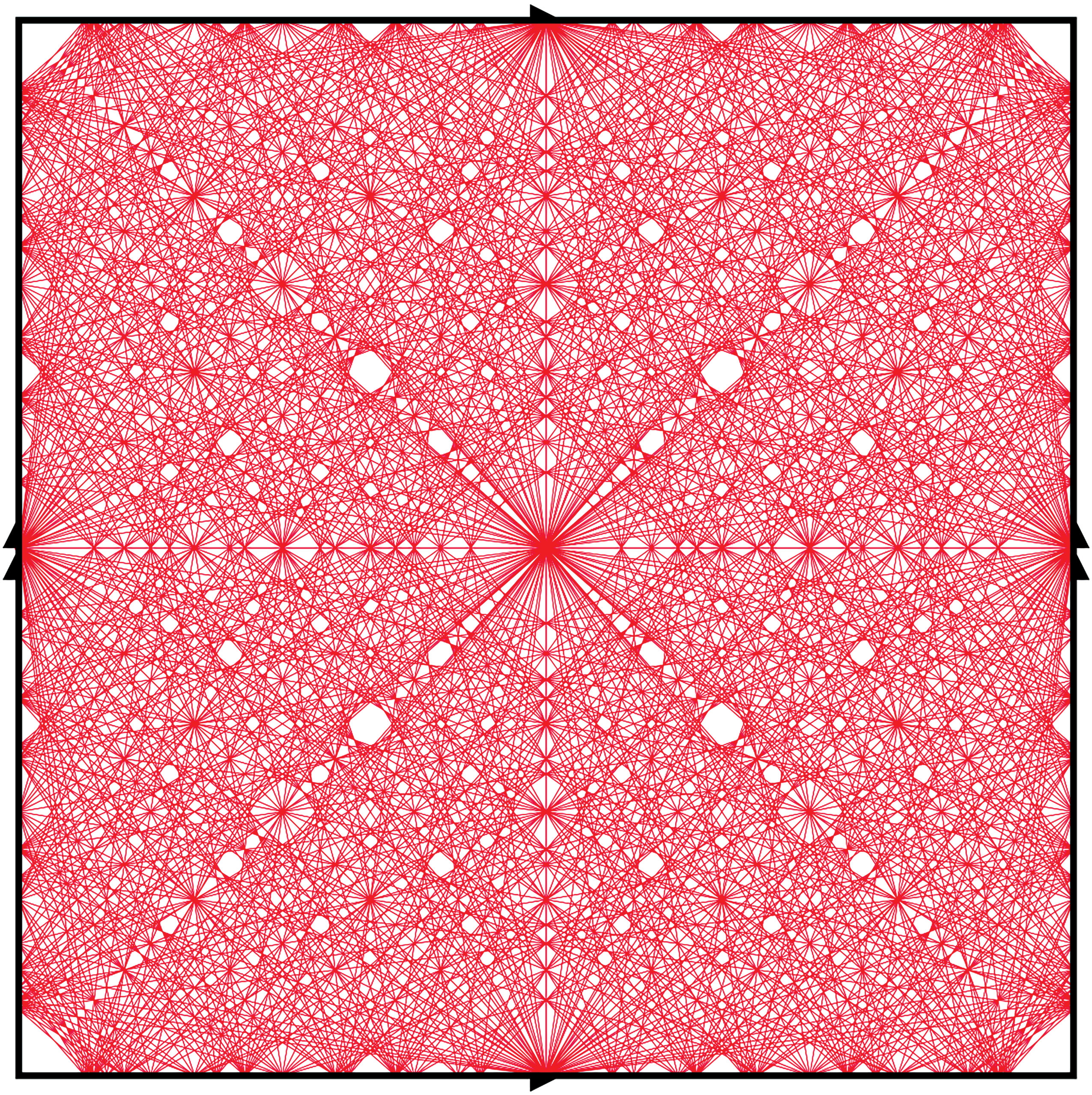
Figure 7 is courtesy of Grace Murray Hopper Collection, Archives Center, National Museum of American History, Smithsonian Institution.

Figure 9 was created by the author using TikZ.

Figures 10 and 11 are courtesy of Manuscripts and Archives, Yale University Library. Reproduction of details from Grace Murray Hopper's PhD dissertation [19] in consultation with family representatives Roger and Deborah Murray.

Author photo was taken by Joe Rabinoff in Salt Lake City, Utah, July 2015. Reproduction is courtesy of Joe Rabinoff.

The Mathematics of Joan Birman



Dan Margalit

Dan Margalit is a Professor in the School of Mathematics at the Georgia Institute of Technology. His email address is `margalit@math.gatech.edu`.

He is supported by the National Science Foundation under Grant No. DMS - 1057874.

Introduction

Joan Birman published her first paper, "On braid groups," in January 1969. That work introduced one of the most important tools in the study of braids and surfaces, now called the Birman exact sequence. Fifty years and more than one hundred papers later, Birman is an active researcher and has long been established as a leading figure in the field of low-dimensional topology.

The goal of this article is to give a broad overview of Birman's mathematics. In the process, we will see several related themes emerge. Time and again, Birman has shown a knack for asking the right questions, for pursuing and embracing unlikely collaborations across mathematical disciplines, and for uncovering and revitalizing hidden or forgotten fields. Because of this, her work has often been ahead of its time, with important implications and applications found years or decades after the original discoveries. For instance, her book on braids is credited with bringing that theory from the fringes to the fore. Similarly, when Birman began working on mapping class groups and Torelli groups, she was working in isolation. Now these are core topics in topology, and her contributions are of fundamental importance. In fact, Birman's work has underpinned two Fields medals.

Birman's research revolves around the theories of knots, braids, mapping class groups of surfaces, and 3-manifolds. Figure 1 shows a diagram of these topics and gives a road map for this article. We will introduce the various objects and the connections between them in the sections indicated. It is a bit of a miracle that these subjects are so closely intertwined. In what follows we will see how Birman's work has influenced and interacted with this beautiful circle of ideas.

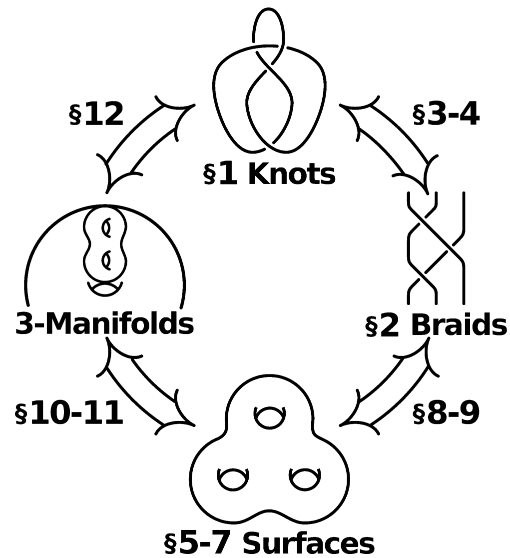


Figure 1. A road map for this article (and Birman's career).

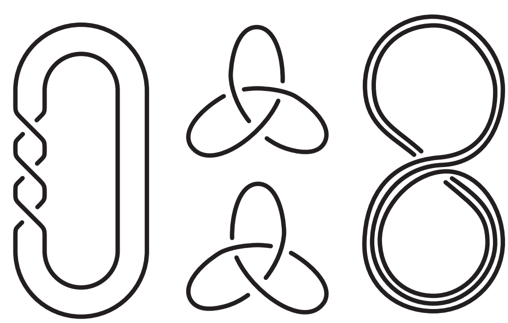


Figure 2. Some examples of knots.

§1 Knots

A knot is the image of a smooth embedding of the circle S^1 into $\mathbb{R}^3$. We can think of a knot as a piece of string with its ends glued together. We can draw a diagram of a knot by projecting it to a plane and indicating the over/undercrossings of the strands by putting a break in the strand that is crossing below; see Figure 2. Two knots are equivalent if they are isotopic, that is, if one knot can be continuously deformed into the other without creating any self-intersections along the way.

The fundamental problem in knot theory is to decide if two knots are equivalent. A (not really) simpler version is to decide if a given knot is equivalent to the trivial knot. The knots in Figure 2 fall into two equivalence classes (left and right trefoils). Which are equivalent?

As this exercise illustrates, knot theory is difficult because there are many diagrams for the same knot that are very different from one another. There is no easy way to move between two different diagrams, and there is no systematic way to choose a canonical diagram for a knot.

Among the many successes of knot theory is the discovery of knot invariants. An invariant for a knot is an object (number, polynomial, etc.) we can associate to a knot with the property that equivalent knots have the same invariant. If we find two knots with different invariants, then they are inequivalent knots.

One of the most famous and important knot invariants is the Alexander polynomial, a Laurent polynomial that can be computed from any knot diagram. The Alexander polynomial is not a complete invariant: it attains the same value on the left- and right-handed trefoil knots, and also Kinoshita and Terasaka found a nontrivial knot with the same Alexander polynomial as the trivial knot. The simplest diagram for the latter has 11 crossings. It is still an open problem to find an easily computable, complete invariant for knots (more on this later).

DOI: https://doi.org/10.1090/noti/1808

Knot theory has applications to statistical mechanics, molecular biology, and chemistry; see Murasugi's book [71] for a survey. Later in this article we will see several connections of knot theory with other parts of topology, group theory, dynamics, and number theory.

§2 Braids

A braid on n strands is a collection of n disjoint paths in $\mathbb{R}^2 \times [0,1]$, connecting n points in $\mathbb{R}^2 \times \{0\}$ to the corresponding points in $\mathbb{R}^2 \times \{1\}$, and intersecting each plane $\mathbb{R}^2 \times \{t\}$ in exactly n points. The n paths are called the strands of the braid.

We consider two braids to be equivalent if they are isotopic, that is, if we can continuously deform one to the other while holding the endpoints fixed and without allowing strands to pass through each other. Figure 3 shows two equivalent braids. The set of braids on n strands forms a group B_n, with the group operation given by stacking braids.

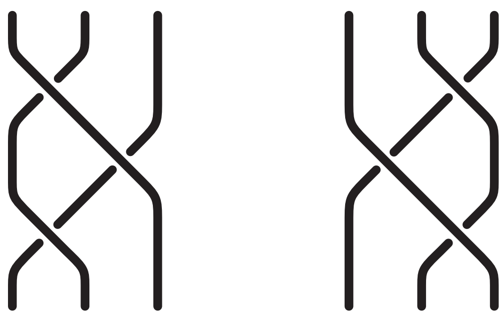


Figure 3. Two equivalent braids.

There is a more succinct (and sophisticated) way to define the braid group. Let C_n denote the configuration space of n distinct points in the plane. We have

$$B_n \cong \pi_1(C_n).$$

The isomorphism is obtained as follows. Let eta be a braid on n strands. For each t in $[0,1]$ we may consider the corresponding plane parallel to the original two planes. If we intersect this plane with the braid eta, we obtain a point in C_n. As t changes from 0 to 1, we obtain a loop in C_n, that is, an element of $\pi_1(C_n)$. This map is the desired isomorphism.

We can now see why the braid group is ubiquitous in mathematics and science: it records the motions of points in the plane. The points can be roots of polynomials, critical values of branched covers, particles in a two-dimensional medium, or autonomous vehicles moving through city streets. See the survey by Birman and her student Brendle for an excellent introduction to the theory [16].

§3 Braids and Knots

There is a simple way to obtain a knot from a braid, namely by connecting the top of the braid to the bottom by n parallel strands. Actually, in general we obtain a link, which is a disjoint union of knots. The resulting knot or link is called a closed braid; see the left-hand side of Figure 4 for an example. In 1923 Alexander proved the remarkable theorem that every knot is equivalent to a closed braid [3].

On the face of it, braids are more tractable than knots because of the group structure, and Alexander's theorem gives us hope of applying our knowledge of braid groups to the theory of knots. The immediate problem is that there are many braids giving rise to the same knot. For instance, if two braids are conjugate, then their braid closures are equivalent.

There are also nonconjugate braids with equivalent closures, and there are braids with different numbers of strands that have equivalent closures. One specific way to construct braids with different numbers of strands and equivalent closures is through stabilization, illustrated in Figure 4. In 1936 Markov announced (without proof) the following surprising theorem: if two braid closures are equivalent, then, up to conjugacy, the braids differ by a finite sequence of stabilizations, destabilizations, and exchange moves (although it was soon realized that the exchange moves were not needed).

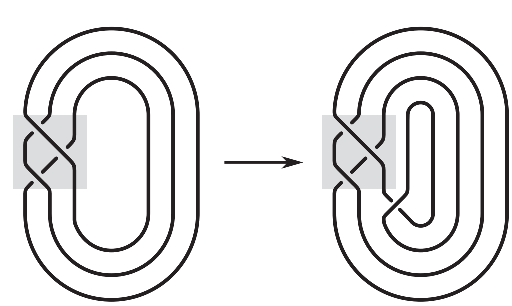


Figure 4. A closed braid and its stabilization.

Four decades later, Birman published a monograph, *Braids, links, and mapping class groups* [12], based on a graduate course she gave at Princeton University during the academic year 1971–72. Her book was the first comprehensive treatment of braid theory, and its appearance represented the birth of the modern theory. It contains in particular the first complete proof of Markov's theorem.

Our discussion of braids and knots so far points us in three natural directions:

1. the *conjugacy problem* for the braid group, namely, the problem of algorithmically determining whether or not two elements of B_n are conjugate;

2. the *algebraic link problem*, namely, the more general problem of algorithmically determining if two braids have equivalent closures; and
3. the big question of whether we can use braid theory to discover new knot invariants.

Birman's monograph focused precisely on these problems. Here we briefly touch on the first two problems, and some contributions to these made by Birman later in her career. In the next section we discuss how the book contributed to the third problem.

With respect to the conjugacy problem, Birman's work has led in two directions. In the 2000s she wrote three papers with Gebhardt and Gonzáles-Meneses [21–23] in which they expand on the Garside approach to the conjugacy problem, explored three decades earlier in Birman's book. A different approach is provided by her paper with Ko and Lee [7]. There, they introduce a new algebraic approach to the braid group, a tool now called the Birman–Ko–Lee monoid for the braid group. This is the second-most cited paper in Birman's catalog.

In the 1990s Birman and Menasco wrote a series of six papers with the title "Studying links via closed braids" [31–36]. The fourth in the series was published in *Inventiones Mathematicae*. A basic question is studied in these papers: If two braids have the same number of strands and have equivalent closures, can we find a sequence of elementary moves that pass from one braid to the other without changing the number of strands? Can we do this algorithmically?

In the end Birman and Menasco did find a "Markov theorem without stabilization," a calculus for dealing with the algebraic link problem [37]. Along the way, they developed connections and applications to the field of contact topology. In particular, they give examples where the isotopy class of a knot and the Bennequin invariant do not fully determine the transverse isotopy class [38]; see also Birman's work with her student Wrinkle [45] as well as the work of Etnyre and Honda [47].

§4 Birman's Book and the Jones Polynomial

While at Princeton, Birman's research focus was on the third problem described in the last section, namely, using braid theory to discover new knot invariants. One tool that becomes available when we have a group in hand is the subject of representation theory. This is relevant to the theory of knot invariants because conjugacy classes of matrices have many natural invariants, such as the determinant.

At the time of Birman's book, only one interesting representation of the braid group was known, namely, the Burau representation. This representation gives a knot invariant as follows: given a knot, choose a braid whose closure is that knot, apply the Burau representation, subtract this matrix from the identity, take the determinant, and then scale by $(1-t)/(1-t^n)$. This conjugacy class invariant for braids interacts nicely with stabilization, and so we indeed obtain a knot invariant.

The knot invariant arising from the Burau representation turns out to be nothing other than the Alexander polynomial. (To paraphrase one of Birman's sayings, when you discover a new knot invariant, your task is to figure out which existing invariant you have just rediscovered.) The Alexander polynomial is of fundamental importance in knot theory, but as mentioned earlier it is not a complete invariant. And without any new representations on the horizon, it seemed hopeless for Birman to use her ideas to extract knot invariants from braids.

But then in 1984, after Birman became a professor at Columbia University, Vaughan Jones asked to meet with Birman to discuss a new representation of the braid group he had discovered through his work on von Neumann algebras. His representation was a direct sum of matrix representations, one of the summands being the Burau representation. From the representation, Jones extracted a conjugacy class invariant for braids. This was not a determinant (as for the Alexander polynomial) but a weighted sum of the traces of the summands [56].

Birman explained the Markov theorem to Jones, who then realized that his conjugacy invariant for braids gave a new invariant of knots, similar to how the Burau representation gives the Alexander polynomial.

Jones' new polynomial was quickly seen to be an improvement over the Alexander polynomial, as it could distinguish the left- and right-handed trefoil knots. Even better, it evaluated nontrivially on the 11-crossing Kinoshita–Terasaka knot [58]. And so the Jones polynomial was born, and a revolution in knot theory was begun.

Jones received the Fields Medal in 1990 for this work. Fittingly, Birman gave the laudation at the International Congress of Mathematicians. See Birman's article from the proceedings [17] and also her personal recollections in this journal [1]. In his *Annals* paper [57], Jones writes, "The author would like to single out Joan Birman among the many recipients of his thanks. Her contribution to this new topic has been of inestimable importance."

Jones showed that his polynomial is not a complete knot invariant: the Conway knot and the 11-crossing Kinoshita–Terasaka knot have the same Jones polynomial. In a paper published in *Inventiones Mathematicae*, Birman further found many inequivalent closed 3-braids with the same Jones polynomial [15]. It is an open question whether or not there is a nontrivial knot with trivial Jones polynomial.

Birman and Wenzl used the theory of the Jones polynomial (specifically, the two-variable polynomial of Kauffman) to construct a new representation of the braid group

[42]. Both Jones and Birman's student Zinno [78] proved that one summand of this representation is the same as the Lawrence representation, famously proved to be faithful by Bigelow [5] and Krammer [61].

Shortly after Jones' discovery, Vassiliev discovered new invariants of knots, named for him (and sometimes called finite-type invariants). Birman and Lin gave a simplified, axiomatic, combinatorial approach to these invariants [29]. This is Birman's most cited paper and was also published in *Inventiones Mathematicae*. Birman wrote a beautiful survey paper explaining this work and the connection to the Jones polynomial [18]; this article won the Chauvenet Prize in 1996.

§5 Mapping Class Groups

We now move on from the world of knots and braids, which are one-dimensional objects, to the realm of surfaces, which are inherently two-dimensional. The theory of mapping class groups of surfaces was initiated by Dehn in the 1920s. Dehn was the doctoral advisor of Magnus who, in turn, was the advisor to Birman. As we will see, mapping class groups will play a prominent role in Birman's career.

To start at the beginning, a surface is a two-dimensional manifold. For each $g \geq 0$ there is a surface S_g of genus g, obtained as the connect sum of g tori (so S_0 is the sphere, and S_1 is the torus). The classification of surfaces says that these are all of the surfaces that are closed (compact and without boundary) and orientable.

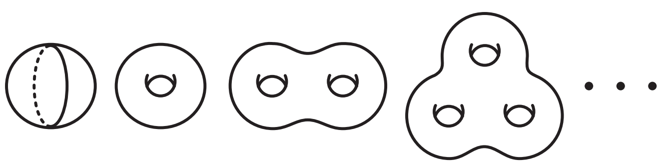


Figure 5. The first few closed, orientable surfaces.

While surfaces are completely classified, there are many open questions, and the theory of surfaces is an active area of research today. Of particular interest is the mapping class group $\mathrm{MCG}(S)$ of a surface S, the group of homotopy classes of homeomorphisms of S. This is a discrete group that encodes the symmetries of S. One source of nontrivial elements of $\mathrm{MCG}(S)$ is the set of rotations of S. For instance the surface S_3 in Figure 5 admits an obvious rotation of order 3.

An important type of infinite order element is a Dehn twist. In Figure 6 we depict a twist of the annulus. A Dehn twist on a surface is a homeomorphism that performs such a twist on some annulus and is the identity on the complement. If c is a simple closed curve in S, then the Dehn twist about an annular neighborhood of c is a well-defined element T_c of $\mathrm{MCG}(S)$.

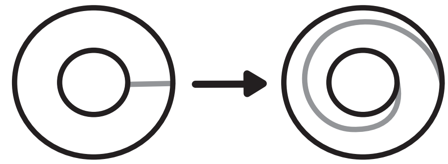


Figure 6. A twist of an annulus.

Dehn proved the foundational theorem that $\mathrm{MCG}(S_g)$ is finitely generated by Dehn twists. Dehn's point of view was motivated by the following analogy:

linear maps : vectors :: mapping classes : curves

More specifically, Dehn was interested in simple closed curves, those with no self-intersections. He referred to the set of these as the arithmetic field of the surface.

After the early work of Dehn and his student Nielsen, the subject of mapping class groups was largely forgotten. Birman reignited interest in the subject through her thesis work (see Section 8, "The Birman Exact Sequence"), her book, and her various survey articles [13, 14, 20]. The subject really exploded with the work of Thurston, which was announced shortly after Birman's book was published; see the next section.

Today, the theory of mapping class groups is a central topic, connected to many fields of mathematics and physics. For instance it can be interpreted as:

1. the outer automorphism group of the fundamental group of the surface;
2. the fundamental group of the moduli space of algebraic curves;
3. the isometry group of Teichmüller space; and
4. the classifying group for surface bundles.

See the primer by Farb and the author [48] for a modern introduction to mapping class groups.

§6 Curves on Surfaces

Birman and Series wrote a number of papers aimed at understanding the nature of the set of simple closed curves in a surface. They gave, for instance, an algorithm for determining if an element of the fundamental group of a surface has a simple representative [39]. They also described a sense in which the action of $\mathrm{MCG}(S)$ on the space of simple closed curves in S is linear, as per Dehn's analogy above [41].

The most influential result of Birman and Series [40] addresses the question, What does the set of simple closed curves look like if we draw them all at once? Precisely, they fix a surface of negative Euler characteristic and a hyperbolic metric on the surface, and they consider the (unique) geodesic representative of each homotopy class of simple closed curves. Their main theorem is that the union of all such geodesics is nowhere dense and has Hausdorff dimension 1.

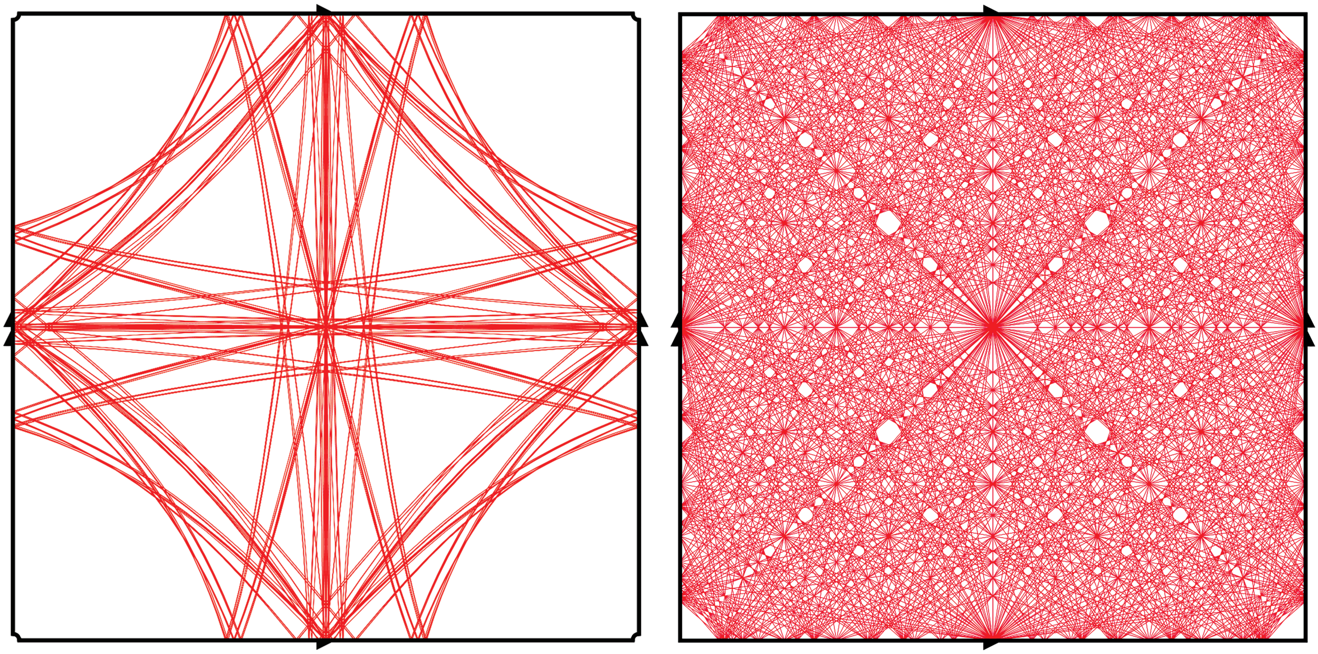


Figure 7. Left(7a): the 88 shortest geodesics on a hyperbolic punctured torus; right(7b): the 88 shortest geodesics on a Euclidean torus.

This result is illustrated by Figure 7. The left side shows a square with the four corners deleted. If we identify opposite sides, we obtain a punctured torus (a torus minus one point). The hyperbolic metric on the latter is mapped to the square by a conformal mapping. Long hyperbolic geodesics are well approximated by arcs of short ones. So even though the picture only shows the 88 shortest simple geodesics, it gives a decent approximation of the union of all simple geodesics.

Here are two striking points of contrast: (1) the union of all closed geodesics (including the ones with self-intersections) is dense; and (2) if we consider a Euclidean torus (the torus obtained by identifying opposite sides of a Euclidean square) and choose one geodesic in each homotopy class of simple closed curves, the resulting union of geodesics is dense (see the right-hand side of Figure 7).

At the end of their paper, Birman and Series suggest another interesting problem: counting the number of simple geodesics as a function of the length. They write:

> In fact the degree of the polynomial $P_0(n)$ bounding the number of simple geodesics of length n is at most $6g+2b-6$, where g is the genus and b the number of boundary components of M... In general the precise nature of the bound seems to be a very interesting number theoretic question.

Many years later, Mirzakhani did find the precise nature of the bound (the upper bound of Birman and Series is also a lower bound), one of the many stunning achievements in her Fields Medal work [68].

The Birman–Series result also plays a central role in the proof of the celebrated McShane identity, which states that for any hyperbolic metric on the punctured torus, we have

$$\sum_{\gamma} \frac{1}{1+e^{\ell(\gamma)}} = 1/2,$$

where the sum is over all simple closed geodesics and $\ell(\gamma)$ denotes the hyperbolic length [66]. This theorem was also generalized by Mirzakhani [67], who used her generalization to compute the volume of moduli space in the Weil–Petersson metric.

§7 Basic Algebraic Properties of the Mapping Class Group

In this section we discuss Birman's work on the following basic algebraic questions about $\mathrm{MCG}(S_g)$:

1. What is the abelianization?
2. What is the rank of a maximal torsion-free abelian subgroup?

These are among the first questions we can ask about any infinite group.

Mumford was one of the few mathematicians who studied the mapping class group in the period between Dehn and Birman. He was interested in the applications to algebraic geometry. What he proved [70] is that any abelian quotient of $\mathrm{MCG}(S_g)$ is a quotient of $\mathbb{Z}/10$ when $g \geq 3$. Birman [11] improved the $\mathbb{Z}/10$ to $\mathbb{Z}/2$. Building on this, her student Powell further improved the $\mathbb{Z}/2$ to the trivial group [73], thus establishing the fundamental theorem that $\mathrm{MCG}(S_g)$ is perfect for $g \geq 3$. This completely answers the first question.

The second question was answered in a joint paper by Birman, Lubotzky, and Birman's student McCarthy [30]. The three of them were working to understand Thurston's groundbreaking work on the mapping class group. As a part of his Fields Medal work, Thurston [75] gave a classification of elements of the mapping class group, now called the Nielsen–Thurston classification. This theorem states that every element of the mapping class group has a representative homeomorphism that preserves a (possibly empty) collection of disjoint curves and, on the complementary pieces, is either of finite order or pseudo-Anosov. A pseudo-Anosov map is one that locally looks like the action of the matrix $\begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}$ on $\mathbb{R}^2$. So there are two invariant foliations, one stretched by λ and one by λ^{-1}.

We should think of Thurston's theorem as a sort of Jordan form for mapping classes. There is one problem: he did not prove that the decomposition along curves was canonical. Birman, Lubotzky, and McCarthy addressed exactly that, by defining the canonical reduction system for a mapping class.

As a result of this work, Birman, Lubotzky, and McCarthy showed that the answer to the second question is $3g-3$ for $\mathrm{MCG}(S_g)$. They further proved that every solvable subgroup of the mapping class group is virtually abelian.

Like the Jordan canonical form for matrices, canonical reduction systems feature prominently in modern theory of mapping class groups, especially in work on their algebraic structure. For instance, Ivanov and McCarthy used canonical reduction systems to prove that mapping class groups satisfy a Tits alternative, thus strengthening the analogy between mapping class groups and arithmetic groups [51, 65].

§8 The Birman Exact Sequence

There are many connections between the theories of braid groups and mapping class groups. The two most important are the Birman exact sequence and the Birman–Hilden theory, discussed in this section and the next. One running theme is that of group presentations for mapping class groups.

Dehn proved that the mapping class group of the torus is isomorphic to $\mathrm{SL}_2(\mathbb{Z})$, which has a well-known finite presentation. In her thesis work, Birman's goal was to find group presentations for other mapping class groups. She succeeded right away in finding an inductive procedure for computing presentations of mapping class groups of surfaces with marked points.

Let S be a surface of negative Euler characteristic, and let $p \in S$. We consider $\mathrm{MCG}(S,p)$, the group of homotopy classes of homeomorphisms of S fixing the point p (it is crucial that the homotopies fix p as well). There is a forgetful map $\mathrm{MCG}(S,p) \to \mathrm{MCG}(S)$. Birman wanted to understand the kernel.

For $[\phi] \in \mathrm{MCG}(S,p)$ to be in the kernel, this means that ϕ is homotopic to the identity as long as we allow p to move during the homotopy. If we follow the path of p throughout this homotopy, we obtain a loop in S, that is, an element of the fundamental group $\pi_1(S,p)$. Birman's theorem is that this identification is well-defined and that it gives an isomorphism of $\pi_1(S,p)$ with the kernel.

The resulting map $\pi_1(S,p) \to \mathrm{MCG}(S,p)$ is usually called the push map because we can think of the image of $\alpha \in \pi_1(S,p)$ as the element of $\mathrm{MCG}(S,p)$ obtained by pushing p along α (Birman originally called this the spin map).

Birman's result is usually stated as saying that the following sequence is exact:

$$1 \to \pi_1(S,p) \to \mathrm{MCG}(S,p) \to \mathrm{MCG}(S) \to 1.$$

Using this, she could promote a presentation of $\mathrm{MCG}(S)$ to a presentation for $\mathrm{MCG}(S,p)$. The Birman exact sequence is ubiquitous in the theory of mapping class groups, as it is used in many inductive arguments.

What is the connection to braid groups? The first step in this direction is to generalize from one point p to a finite set of points $P = \{p_1, \dots, p_n\}$. The group $\mathrm{MCG}(S,P)$ is the group of homotopy classes of homeomorphisms of S fixing P as a set. Let $C_n(S)$ denote the space of configurations of n distinct points in S. Birman's more general exact sequence is

$$1 \to \pi_1(C_n(S),P) \to \mathrm{MCG}(S,P) \to \mathrm{MCG}(S) \to 1.$$

When $n=1$, the space $C_n(S)$ is homeomorphic to S, and so we obtain the first exact sequence above. Recall that B_n is defined as $\pi_1(C_n(\mathbb{R}^2),P)$. The group $\pi_1(C_n(S),P)$ is known as a surface braid group. We can visualize the elements as braided strands in $S \times [0,1]$. As a special case, when S is the disk, we conclude that B_n is isomorphic to the mapping class group of a disk with n marked points.

Birman used the more general exact sequence in her thesis to obtain presentations for the mapping class groups of the torus with any number of marked points [10]. The surface of genus 2 would have to wait for her work with Hilden.

§9 The Birman–Hilden Theory

After graduating from New York University's Courant Institute in 1968, Birman took a job at Stevens Institute of Technology, where she began a very successful collaboration with Hilden, a graduate student there at the time.

Birman and Hilden originally set out to find a presentation for $\mathrm{MCG}(S_2)$, the next natural mountain to climb. The key idea in their work is to relate $\mathrm{MCG}(S_2)$ to a braid group in the following way. The hyperelliptic involution

$\iota : S_2 \to S_2$ is the rotation by π about the axis indicated in Figure 8.

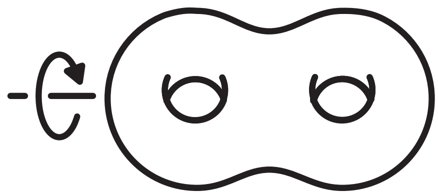


Figure 8. The hyperelliptic involution of S_2.

The quotient $S_2/\langle \iota \rangle$ is a sphere $S_{0,6}$ with six distinguished points (the images of the six fixed points of ι). Birman and Hilden proved that there is an isomorphism

$$\mathrm{MCG}(S_2)/\langle [\iota] \rangle \xrightarrow{\cong} \mathrm{MCG}(S_{0,6}).$$

Since $\mathrm{MCG}(S_{0,6})$ is closely related to a braid group (with the sphere replacing the disk), this allowed them to convert a known presentation for $\mathrm{MCG}(S_{0,6})$ into a presentation for $\mathrm{MCG}(S_2)$. This work is the subject of Birman's article, "My favorite paper" [9].

The above isomorphism is defined as follows. As observed earlier by Birman, every element of $\mathrm{MCG}(S_2)$ has a representative that commutes with ι. Such a representative descends to a homeomorphism of $S_{0,6}$ and hence gives an element of $\mathrm{MCG}(S_{0,6})$. The hard part of their theorem is showing that this map is well-defined, that is, that it interacts well with homotopies.

Birman and Hilden vastly generalized this theorem in a series of papers on hyperelliptic and symmetric mapping class groups [24–26], culminating in their most general result [28], which was published in *Annals of Mathematics*. This work was later generalized by MacLachlan and Harvey [63] and by Winarski [76], who gave Teichmüller-theoretic and combinatorial-topological points of view.

The Birman–Hilden theory gives a dictionary between the theories of braid groups and mapping class groups, with important applications on both sides. For instance it is used in the proof that $\mathrm{MCG}(S_2)$ is linear [6, 60] and also in the resolution of a question of Magnus about the action of the braid group on the fundamental group of the punctured disk [28]. We refer the reader to our survey with Winarski for a detailed discussion [64].

§10 Heegaard Splittings, Torelli Groups, and Homology Spheres

We now turn to the interface between the theories of surfaces and 3-manifolds. A 3-manifold is the three-dimensional analogue of a surface, that is, a space that locally looks like $\mathbb{R}^3$. A first example is the 3-sphere S^3. We can use stereographic projection to identify S^3 as $\mathbb{R}^3$ with one added point at infinity, in much the same way that we identify S^2 as $\mathbb{R}^2$ with a point at infinity.

In this section we will focus on one particular construction of 3-manifolds from surfaces, namely Heegaard splittings. If S_g is the surface of a donut with g donut holes, then the handlebody H_g is the donut itself. By gluing two copies of H_g along their boundaries, we obtain a 3-manifold without boundary. For each g there is a particular gluing $\psi : S_g \to S_g$ that results in the sphere S^3. (The usual embedding of H_g in $\mathbb{R}^3 \subseteq S^3$ is a realization of this gluing: the outside of H_g is another copy of H_g!) In general, the decomposition of a 3-manifold into two handlebodies glued along their boundary is called a Heegaard splitting.

If we take any homeomorphism ϕ of S_g and post-compose the gluing map ψ by ϕ, we obtain a new 3-manifold. The resulting 3-manifold only depends on the mapping class $[\phi] \in \mathrm{MCG}(S_g)$. What is more, every closed, orientable 3-manifold arises in this way. The upshot is that the theory of Heegaard splittings gives us a set map

$$\mathrm{MCG}(S_g) \to \text{3-manifolds}.$$

The mapping class group $\mathrm{MCG}(S_g)$ acts on the first homology group $H_1(S_g)$. The kernel of this action is called the Torelli group $\mathcal{I}(S_g)$. By the Mayer–Vietoris theorem, we have the restriction

$$\mathcal{I}(S_g) \to \text{homology 3-spheres}.$$

Here, a homology 3-sphere is a 3-manifold that has the same homology groups as S^3. This is an important subclass of 3-manifolds. Indeed, the fact that there exist nontrivial homology 3-spheres is the reason that the Poincaré conjecture cannot be stated in terms of homology alone (and this is what forced Poincaré to invent π_1).

Birman published a number of works on Heegaard splittings, specifically with the aim of classifying 3-manifolds through the lens of the mapping class group. For instance, with Hilden [27] she gave an algorithm to determine if a manifold with a given Heegaard splitting is homeomorphic to S^3.

§11 Birman's Work on Torelli Groups

Birman made two monumental contributions to the theory of Torelli groups. In particular, her work was aimed at the following questions:

1. What is a natural generating set for the Torelli group?
2. What are the abelian quotients of the Torelli group?
3. Is the Torelli group finitely generated?

As with mapping class groups, these are among the first properties we would like to know about a group.

There is also a connection with algebraic geometry: the Torelli group encodes the fundamental group of the Torelli space, the space of framed curves of genus g. The period mapping takes this space to the Siegel upper half-space, sending a framed curve to its period matrix. (Torelli is the name of an Italian algebraic geometer.) As such, the above questions can be reinterpreted as basic questions about the topology of Torelli space.

Birman spent the academic year 1969–70 in Paris. By her own account, she was mathematically isolated there and discouraged [1]. But she had an idea for how to attack the first question by brute-force calculation. The starting point is that the mapping class group $\mathrm{MCG}(S_g)$ and the Torelli group $\mathcal{I}(S_g)$ fit into a short exact sequence

$$1 \to \mathcal{I}(S_g) \to \mathrm{MCG}(S_g) \to \mathrm{Sp}_{2g}(\mathbb{Z}) \to 1.$$

The group $\mathrm{Sp}_{2g}(\mathbb{Z})$ is isomorphic to the automorphism group of $H_1(S_g;\mathbb{Z}) \cong \mathbb{Z}^{2g}$; we have the symplectic group here instead of the whole general linear group because automorphisms preserve the algebraic intersection form, which is symplectic. From this point of view, we can think of $\mathrm{Sp}_{2g}(\mathbb{Z})$ as capturing the linear, easy-to-understand aspects of $\mathrm{MCG}(S_g)$ and of $\mathcal{I}(S_g)$ as encapsulating the more difficult, mysterious aspects.

Birman knew that the defining relations for $\mathrm{Sp}_{2g}(\mathbb{Z})$ correspond to generators for $\mathcal{I}(S_g)$ (this is a general principle that applies to any short exact sequence of groups). So the task then was to find a reasonable group presentation for $\mathrm{Sp}_{2g}(\mathbb{Z})$. She succeeded and obtained a presentation with three families of generators and 10 families of relations.

Birman's student Powell then gave simple descriptions of the resulting generators for $\mathcal{I}(S_g)$: they are Dehn twists about separating curves and bounding pair maps [73]. A bounding pair map is $T_a T_b^{-1}$, where a and b are disjoint, homologous, nonseparating curves; see Figure 9. Putman, who gave a geometric proof of the Birman–Powell result in his thesis [74], describes Birman's work as "absolutely heroic."

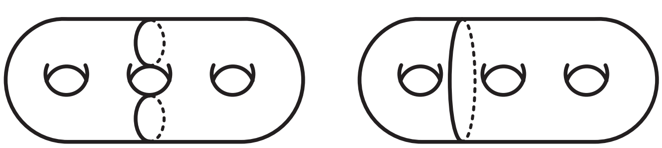


Figure 9. Left: a bounding pair; right: a separating curve.

Birman and Craggs took aim at the second and third questions, and they made a most spectacular contribution. They showed that, unlike $\mathrm{MCG}(S_g)$, the group $\mathcal{I}(S_g)$ does have nontrivial abelian quotients. They found a family of homomorphisms $\rho_\psi : \mathcal{I}(S_g) \to \mathbb{Z}/2$. Surprisingly, the definition involves the theories of 3- and 4-manifolds. One hope they had was that there would be infinitely many distinct such homomorphisms, thus proving that $\mathcal{I}(S_g)$ was not finitely generated.

In order to specify one of the Birman–Craggs homomorphisms, we need to fix some Heegaard splitting ψ of S^3. Now let $f \in \mathcal{I}(S_g)$. As in Section 10, "Heegaard Splittings, Torelli Groups, and Homology Spheres," f determines a homology 3-sphere M_f. Every homology 3-sphere is the boundary of some 4-manifold. The Rokhlin invariant of M_f is the signature of this 4-manifold, divided by 8, mod 2 (by Rokhlin's theorem, this is well-defined). This element of $\mathbb{Z}/2$ is $\rho_\psi(f)$. Miraculously, this defines a homomorphism $\mathcal{I}(S_g) \to \mathbb{Z}/2$. The proof features what is probably the first instance of a 4-manifold trisection, a tool popularized four decades later by David Gay and Robion Kirby [79].

Several years after these works, Johnson arrived on the scene. In a stunning series of deep, beautiful papers, he expanded on the work of Birman and her collaborators. He proved [55] that $\mathcal{I}(S_g)$ is finitely generated for $g \geq 3$. Also he classified the Birman–Craggs homomorphisms—showing directly that there were only finitely many—and gave a complete description of the abelianization of $\mathcal{I}(S_g)$ [52]. (Amazingly, there is still no definition of these homomorphisms that does not involve the construction of a 4-manifold.) As a byproduct, Johnson showed that $\mathcal{I}(S_g)$ cannot be generated by Dehn twists about separating curves, disproving a conjecture of Birman.

See Johnson's delightful survey for more about his work [53]. In the survey, Johnson notes that the interest in Torelli groups from topologists "was initiated principally through the work of Joan Birman" [54].

§12 Lorenz Knots

We end by discussing the work of Birman and Williams on Lorenz knots in the early 1980s. This is a fitting finale, as it combines all four of the main objects of study in this article. It is also a prime example of work that was ahead of its time, with 94 of its 106 citations on MathSciNet® coming after the year 2000.

E. N. Lorenz was a pioneer of chaos theory. He was particularly interested in the weather, and whether it was deterministic. Lorenz is perhaps most famous for coining the phrase "butterfly effect."

In order to help understand weather patterns, Lorenz devised a simplified version of the Navier–Stokes equations, a system of three ordinary differential equations in three variables [62]. This system has a strange attractor, called the Lorenz attractor, shown in the top of Figure 10. Forward trajectories of points converge to the attractor and, once there, stay forever.

A Lorenz knot is a knot obtained as a periodic orbit in the Lorenz attractor. Williams showed that Lorenz knots are exactly the ones that can be drawn on the "template" shown at the bottom of Figure 10.

A Lorenz braid is a braid consisting of strands that either go monotonically left to right or from right to left, where

Figure 10. Top(10a): the Lorenz attractor; bottom(10b): the Lorenz template.

the strands going from left to right pass over the strands going from right to left, and where neither the left-to-right nor the right-to-left strands cross amongst themselves; see Figure 11. Lorenz knots can also be described as the closures of Lorenz braids.

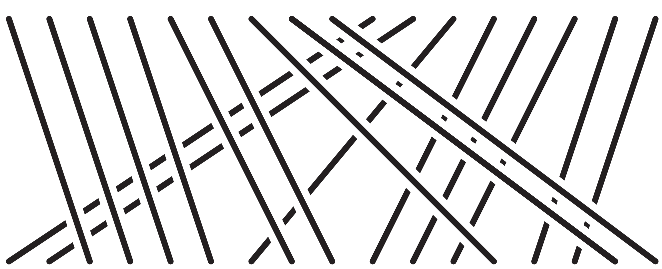


Figure 11. A Lorenz braid.

Williams approached Birman at a conference and asked her if she could identify some of the knots he was studying. She could, and their discussion quickly turned into a fruitful collaboration. In their first paper [44], Birman and Williams proved many theorems about Lorenz knots, including:

1. There are infinitely many (inequivalent) Lorenz knots.
2. Lorenz knots are prime.
3. Every algebraic knot is a Lorenz knot.
4. Every Lorenz knot is fibered.

In the third theorem, an algebraic knot is any component of the link of an isolated singularity of a complex curve. The fourth theorem requires some explanation. We can construct a 3-manifold from a surface S by the mapping torus construction: for $[\phi] \in \mathrm{MCG}(S)$, we take the product $S \times [0,1]$ and glue $S \times \{0\}$ to $S \times \{1\}$ by ϕ. The resulting 3-manifold has a natural map to S^1 with fiber S, and we say that the 3-manifold is fibered. A knot in $\mathbb{R}^3$ is said to be fibered if its complement in S^3 is a fibered 3-manifold.

Two decades after Birman and Williams, Ghys entered the picture. He was studying the manifold $M = \mathrm{PSL}_2(\mathbb{R})/\mathrm{PSL}_2(\mathbb{Z})$. The manifold M is homeomorphic to the complement in S^3 of the trefoil knot, and it can also be described as the unit tangent bundle of the modular surface (the quotient of the hyperbolic plane by $\mathrm{PSL}_2(\mathbb{Z})$). From the latter description, M has a geodesic flow. Ghys was studying the closed orbits in this flow, and he proved that the knots arising from these closed orbits are in natural bijection with the Lorenz knots (the connection was further investigated by Pinsky [72]). He further showed that the Rademacher function exactly records the linking number of each knot with the missing trefoil. We recommend Ghys's beautiful survey, written on the occasion of his plenary lecture at the International Congress of Mathematicians [50].

We next turn to the question, How common are Lorenz knots? Dehornoy, Ghys, and Jablon showed that of the 1,701,936 knots with at most 16 crossings in their diagrams, only 20 are Lorenz knots. And so from this point of view they appear to be rather rare. Birman and her postdoc Kofman took a different point of view. In order to explain it, we take a detour into hyperbolic geometry and the classification of 3-manifolds.

Thurston revolutionized the theory of 3-manifolds by showing that many knots are hyperbolic; that is, their complements in S^3 could be given complete Riemannian metrics of constant sectional curvature -1. By the Mostow rigidity theorem, hyperbolic structures on 3-manifolds are unique. In particular, a hyperbolic knot has a well-defined volume.

Thurston's work on knots eventually led him to formulate his geometrization conjecture, which shaped the field for several decades. The conjecture states that every 3-manifold can be decomposed into geometric pieces, namely, Seifert-fibered spaces (completely classified in the 1930s by Seifert) and hyperbolic manifolds. The Poincaré conjecture is a special case of Thurston's conjecture because there are no counterexamples to the latter among the Seifert-fiber -ed spaces or the closed hyperbolic manifolds (which have infinite fundamental group).

The geometrization conjecture was famously proved by Perelman in 2003; see [46, 59, 69]. More recently, Agol and Wise proved that every closed hyperbolic 3-manifold has a finite cover that is fibered, verifying another conjecture of Thurston [2, 4, 77]. This gives a satisfying description of the hyperbolic pieces of a 3-manifold: up to taking finite covers, they all come from surface homeomorphisms.

We return now to our story about Lorenz knots. Rather than organizing knots by the number of crossings in their diagrams, Birman and Kofman organized the hyperbolic knots by their volumes. They showed that of the 201 hyperbolic knots of smallest volume, more than half of them are Lorenz knots [8]. So among all knots, Lorenz knots are extremely rare, but among the small-volume hyperbolic knots, Lorenz knots are quite prevalent.

Birman and Williams wrote a companion paper [43] where they studied a different flow on S^3 and discovered an appropriate template in that case as well. In his gem of a thesis, Ghrist [49] showed that this flow is universal, in that it contains *all* knots as closed orbits, disproving a conjecture of Birman and Williams.

There are many other intriguing aspects to the story and tantalizing questions to answer. As Birman writes at the end of her survey [19], "There is a big world out there, and a great deal of structure, waiting to be discovered!"

Epilogue

A distinguishing feature of Birman's career is that her research has been motivated by her own vision, interests, and curiosity. There are very few instances where Birman was trying to answer someone else's question or solve someone else's problem. While this may seem like a risky way to approach a career in mathematics, it is hard to argue with the results. Besides the beautiful mathematics she has produced by herself and with her collaborators, she has had (as we have seen) a direct impact on two Fields Medals (Jones' and Mirzakhani's) and a plenary address at the International Congress of Mathematicians (Ghys'), among the many works she has helped to inspire.

As we touched on at the outset and throughout this article, Birman's work was in many cases ahead of its time, her foundational work finding applications (and appreciation) many years after the original discovery. Braid groups, mapping class groups, Torelli groups, and Lorenz knots were fringe topics when she started. With the breakthroughs of Jones, Mirzakhani, Thurston, Johnson, and Ghys we have seen the impact and validation of Birman's work.

As a recent collaborator of Birman's and as a researcher in the same field, the author has had the pleasure of seeing Birman's mathematics from up close and being inspired by her work. We eagerly look forward to the next chapters of Birman's career, including new discoveries by Birman herself and new perspectives on her prior work, yet to be uncovered.

References

[1] Interview with Joan Birman. *Notices of the American Mathematical Society*, 54(1):20–29, 2007. MR2275922

[2] Agol I. The virtual Haken conjecture. *Doc. Math.*, 18:1045–1087, 2013. With an appendix by Agol, Daniel Groves, and Jason Manning. MR3104553

[3] Alexander J. A lemma on systems of knotted curves. *Proc. Natl. Acad. Sci. USA*, 9:93–95, 1923.

[4] Bestvina M. Geometric group theory and 3-manifolds hand in hand: the fulfillment of Thurston's vision. *Bull. Amer. Math. Soc. (N.S.)*, 51(1):53–70, 2014. MR3119822

[5] Bigelow SJ. Braid groups are linear. *J. Amer. Math. Soc.*, 14(2):471–486, 2001. MR1815219

[6] Bigelow SJ and Budney RD. The mapping class group of a genus two surface is linear. *Algebr. Geom. Topol.*, 1:699–708, 2001. MR1875613

[7] Birman J, Ko KH, and Lee SJ. A new approach to the word and conjugacy problems in the braid groups. *Adv. Math.*, 139(2):322–353, 1998. MR1654165

[8] Birman J and Kofman I. A new twist on Lorenz links. *J. Topol.*, 2(2):227–248, 2009. MR2529294

[9] Birman JS. My favorite paper. *Celebratio Mathematica*, (to appear). MR1873103

[10] Birman JS. Mapping class groups and their relationship to braid groups. *Comm. Pure Appl. Math.*, 22:213–238, 1969. MR0243519

[11] Birman JS. Abelian quotients of the mapping class group of a 2-manifold. *Bull. Amer. Math. Soc.*, 76:147–150, 1970. MR0249603

[12] Birman JS. *Braids, links, and mapping class groups*. Annals of Mathematics Studies, No. 82, Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, 1974. MR0375281

[13] Birman JS. Mapping class groups of surfaces: a survey. Discontinuous groups and Riemann surfaces (Proc. Conf., Univ. Maryland, College Park, Md., 1973), pp. 57–71. Ann. of Math. Studies, No. 79, 1974. MR0380762

[14] Birman JS. The algebraic structure of surface mapping class groups. *Discrete groups and automorphic functions (Proc. Conf., Cambridge, 1975)*, pp. 163–198, 1977. MR0488019

[15] Birman JS. On the Jones polynomial of closed 3-braids. *Invent. Math.*, 81(2):287–294, 1985. MR799267

[16] Birman JS. Mapping class groups of surfaces. In Braids (Santa Cruz, CA, 1986), pp. 13–43, *Contemp. Math.*, 78, Amer. Math. Soc., Providence, RI, 1988. MR975076

[17] Birman JS. The work of Vaughan F. R. Jones. In *Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990)*, 9–18. Math. Soc. Japan, Tokyo, 1991. MR1159199

[18] Birman JS. New points of view in knot theory. *Bull. Amer. Math. Soc. (N.S.)*, 28(2):253–287, 1993. MR1191478

[19] Birman JS. The mathematics of Lorenz knots. In Topology and dynamics of chaos, pp. 127–148, *World Sci. Ser. Nonlinear Sci. Ser. A Monogr. Treatises*, 84, World Sci. Publ., Hackensack, NJ, 2013. MR3289735

[20] Birman JS and Brendle TE. Braids: a survey. In *Handbook of knot theory*, pp. 19–103. Elsevier B. V., Amsterdam, 2005. MR2179260

[21] Birman JS, Gebhardt V, and González-Meneses J. Conjugacy in Garside groups. I. Cyclings, powers and rigidity. *Groups Geom. Dyn.*, 1(3):221–279, 2007. MR2314045

[22] Birman JS, Gebhardt V, and González-Meneses J. Conjugacy in Garside groups. III. Periodic braids. *J. Algebra*, 316(2):746–776, 2007. MR2358613

[23] Birman JS, Gebhardt V, and González-Meneses J. Conjugacy in Garside groups. II. Structure of the ultra summit set. *Groups Geom. Dyn.*, 2(1):13–61, 2008. MR2367207

[24] Birman JS and Hilden HM. On the mapping class groups of closed surfaces as covering spaces. Advances in the theory of Riemann surfaces (Proc. Conf., Stony Brook, NY, 1969) pp. 81–115. Ann. of Math. Studies, No. 66, Princeton Univ. Press, Princeton, NJ, 1971. MR0292082

[25] Birman JS and Hilden HM. Isotopies of homeomorphisms of Riemann surfaces and a theorem about Artin's braid group. *Bull. Amer. Math. Soc.*, 78:1002–1004, 1972. MR0307217

[26] Birman JS and Hilden HM. Lifting and projecting homeomorphisms. *Arch. Math. (Basel)*, 23:428–434, 1972. MR0321071

[27] Birman JS and Hilden HM. The homeomorphism problem for S^3. *Bull. Amer. Math. Soc.*, 79:1006–1010, 1973. MR0319180

[28] Birman JS and Hilden HM. On isotopies of homeomorphisms of Riemann surfaces. *Ann. of Math. (2)*, 97:424–439, 1973. MR0325959

[29] Birman JS and Lin X-S. Knot polynomials and Vassiliev's invariants. *Invent. Math.*, 111(2):225–270, 1993. MR1198809

[30] Birman JS, Lubotzky A, and McCarthy J. Abelian and solvable subgroups of the mapping class groups. *Duke Math. J.*, 50(4):1107–1120, 1983. MR726319

[31] Birman JS and Menasco WW. Studying links via closed braids. IV. Composite links and split links. *Invent. Math.*, 102(1):115–139, 1990. MR1069243

[32] Birman JS and Menasco WW. Studying links via closed braids. II. On a theorem of Bennequin. *Topology Appl.*, 40(1):71–82, 1991. MR1114092

[33] Birman JS and Menasco WW. Studying links via closed braids. I. A finiteness theorem. *Pacific J. Math.*, 154(1):17–36, 1992. MR1154731

[34] Birman JS and Menasco WW. Studying links via closed braids. V. The unlink. *Trans. Amer. Math. Soc.*, 329(2):585–606, 1992. MR1030509

[35] Birman JS and Menasco WW. Studying links via closed braids. VI. A nonfiniteness theorem. *Pacific J. Math.*, 156(2):265–285, 1992. MR1186805

[36] Birman JS and Menasco WW. Studying links via closed braids. III. Classifying links which are closed 3-braids. *Pacific J. Math.*, 161(1):25–113, 1993. MR1237139

[37] Birman JS and Menasco WW. Stabilization in the braid groups. I. MTWS. *Geom. Topol.*, 10:413–540, 2006. MR2224463

[38] Birman JS and Menasco WW. Stabilization in the braid groups. II. Transversal simplicity of knots. *Geom. Topol.*, 10:1425–1452, 2006. MR2255503

[39] Birman JS and Series C. An algorithm for simple curves on surfaces. *J. London Math. Soc. (2)*, 29(2):331–342, 1984. MR744104

[40] Birman JS and Series C. Geodesics with bounded intersection number on surfaces are sparsely distributed. *Topology*, 24(2):217–225, 1985. MR793185

[41] Birman JS and Series C. Algebraic linearity for an automorphism of a surface group. *J. Pure Appl. Algebra*, 52(3):227–275, 1988. MR952081

[42] Birman JS and Wenzl H. Braids, link polynomials and a new algebra. *Trans. Amer. Math. Soc.*, 313(1):249–273, 1989. MR992598

[43] Birman JS and Williams RF. Knotted periodic orbits in dynamical system. II. Knot holders for fibered knots. In Low-dimensional topology (San Francisco, Calif., 1981), pp. 1–60 *Contemp. Math.*, 120, Amer. Math. Soc., Providence, RI, 1983. MR718132

[44] Birman JS and Williams RF. Knotted periodic orbits in dynamical systems. I. Lorenz's equations. *Topology*, 22(1):47–82, 1983. MR682059

[45] Birman JS and Wrinkle NC. On transversally simple knots. *J. Differential Geom.*, 55(2):325–354, 2000. MR1847313

[46] Cao H-D and Zhu X-P. A complete proof of the Poincaré and geometrization conjectures—application of the Hamilton–Perelman theory of the Ricci flow. *Asian J. Math.*, 10(2):165–492, 2006. MR2233789

[47] Etnyre JB and Honda K. Cabling and transverse simplicity. *Ann. of Math. (2)*, 162(3):1305–1333, 2005. MR2179731

[48] Farb B and Margalit D. *A Primer on Mapping Class Groups*. Princeton Mathematical Series, 49. Princeton University Press, 2012. MR2850125

[49] Ghrist RW. Flows on S^3 supporting all links as orbits. *Electron. Res. Announc. Amer. Math. Soc.*, 1(2):91–97, 1995. MR1350685

[50] Ghys E. Knots and dynamics. In *International Congress of Mathematicians. Vol. I*, 247–277. Eur. Math. Soc., Zürich, 2007. MR2334193

[51] Ivanov NV. Algebraic properties of the Teichmüller modular group. *Dokl. Akad. Nauk SSSR*, 275(4):786–789, 1984. MR745513

[52] Johnson D. The structure of the Torelli group. I. A finite set of generators for $\mathcal{I}$. *Ann. of Math. (2)*, 118(3):423–442, 1983. MR727699

[53] Johnson D. A survey of the Torelli group. In Low-dimensional topology (San Francisco, Calif., 1981), pp. 165–179, *Contemp. Math.*, 20, Amer. Math. Soc., Providence, RI, 1983. MR718141

[54] Johnson D. The structure of the Torelli group. II. A chracterization of the group generated by twists on bounding curves.*Topology* 24(2):113–126, 1985. MR0793178

[55] Johnson D. The structure of the Torelli group. III. The abelianization of $\mathcal{T}$. *Topology*, 24(2):127–144, 1985. MR793179

[56] Jones VFR. Index for subfactors. *Invent. Math.*, 72(1):1–25, 1983. MR696688

[57] Jones VFR. Hecke algebra representations of braid groups and link polynomials. *Ann. of Math. (2)*, 126(2):335–388, 1987. MR908150
[58] Jones VFR. A polynomial invariant for knots via von Neumann algebras. *Bull. Amer. Math. Soc. (N.S.)*, 12(1):103–111, 1985. MR766964
[59] Kleiner B and Lott J. Notes on Perelman's papers. *Geometry & Topology*, 12(5):2587–2855, Nov 2008. MR2460872
[60] Korkmaz M. On the linearity of certain mapping class groups. *Turkish J. Math.*, 24(4):367–371, 2000. MR1803819
[61] Krammer D. Braid groups are linear. *Ann. of Math. (2)*, 155(1):131–156, 2002. MR1888796
[62] Lorenz EN. Deterministic, non-periodic flows. *J. Atmos. Sci.*, 20:130–141, 1963.
[63] Maclachlan C and Harvey WJ. On mapping-class groups and Teichmüller spaces. *Proc. London Math. Soc. (3)*, 30(part 4):496–512, 1975. MR0374414
[64] Margalit D and Winarski R. The Birman–Hilden theory. preprint.
[65] McCarthy J. A "Tits-alternative" for subgroups of surface mapping class groups. *Trans. Amer. Math. Soc.*, 291(2):583–612, 1985. MR800253
[66] McShane G. Simple geodesics and a series constant over Teichmuller space. *Invent. Math.*, 132(3):607–632, 1998. MR1625712
[67] Mirzakhani M. Weil–Petersson volumes and intersection theory on the moduli space of curves. *J. Amer. Math. Soc.*, 20(1):1–23, 2007. MR2257394
[68] Mirzakhani M. Growth of the number of simple closed geodesics on hyperbolic surfaces. *Ann. of Math. (2)*, 168(1):97–125, 2008. MR2415399
[69] Morgan J and Tian G. *Ricci flow and the Poincaré conjecture*, Clay Mathematics Monographs 3, American Mathematical Society, Providence, RI; Clay Mathematics Institute, Cambridge, MA, 2007. MR2334563
[70] Mumford D. Abelian quotients of the Teichmüller modular group. *J. Analyse Math.*, 18:227–244, 1967. MR0219543
[71] Murasugi K. *Knot theory and its applications*. Birkhäuser Boston, Inc., Boston, MA, 1996. Translated from the 1993 Japanese original by Bohdan Kurpita. MR1391727
[72] Pinsky T. On the topology of the Lorenz system. *Proc. A.*, 473(2205):20170374, 11, 2017. MR3710339
[73] Powell J. Two theorems on the mapping class group of a surface. *Proc. Amer. Math. Soc.*, 68(3):347–350, 1978. MR0494115
[74] Putman A. Cutting and pasting in the Torelli group. *Geom. Topol.*, 11:829–865, 2007. MR2302503
[75] Thurston WP. On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Amer. Math. Soc. (N.S.)*, 19(2):417–431, 1988. MR956596
[76] Winarski RR. Symmetry, isotopy, and irregular covers. *Geom. Dedicata*, 177:213–227, 2015. MR3370031
[77] Wise DT. *From riches to raags: 3-manifolds, right-angled Artin groups, and cubical geometry*, *CBMS Regional Conference Series in Mathematics*, 117. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 2012. MR2986461
[78] Zinno MG. On Krammer's representation of the braid group. *Math. Ann.*, 321(1):197–211, 2001. MR1857374
[79] Gay D and Kirby R. Trisecting 4-manifolds. *Geom. Topol.*, 20 (6):3097–3132, 2016. MR3590351

ACKNOWLEDGMENT. We would like to thank Joan Birman for her help in preparing this article, Shane Scott for making the figures, and Tara Brendle, Lei Chen, Benson Farb, Vaughan Jones, Justin Lanier, Kevin Wortman, and an anonymous referee for helpful conversations.

Dan Margalit

Credits

The opening page figure is by David Dumas (see dumas.io/birmanseries/).
Figures 1-6; 8-9; 10b and 11 are by Shane Scott.
Figure 7 is by David Dumas (see dumas.io/birmanseries/).
Figure 10a is by Thierry Dugnolle [CC BY-SA 4.0 (creativecommons.org/licenses/by-sa/4.0)], from Wikimedia Commons.
Author photo courtesy of Joseph Rabinoff.

In Memory of Marina Ratner 1938–2017

DOI: http://dx.doi.org/10.1090/noti1828

MEMORIAL TRIBUTE

Marina Ratner

S. G. Dani

This article has been reproduced with permission from *Bhāvanā*, October 2017, Vol. 1, Issue 4. See bhavana.org.in/marina-ratner-1938-2017.

S. G. Dani was affiliated with the Tata Institute of Fundamental Research (TIFR), Mumbai for over four decades, until mandatory retirement in 2012. Subsequently, he was associated with IIT Bombay and, more recently, is with the UM-DAE Centre for Excellence in Basic Sciences, Mumbai, a collaborative endeavor of the University of Mumbai and the Department of Atomic Emergy of the Government of India. His email address is shrigodani@gmail.com.

DOI: http://dx.doi.org/10.1090/noti1835

Marina Evseevna Ratner, renowned for her work in dynamics, passed away on July 7, 2017, at her home at El Cerrito, California, USA, at the age of 78. Her profound contributions, establishing the Raghunathan conjecture and its variants, from the 1990s when she was in her early fifties, have become a milestone in homogeneous dynamics and have had an impact on the study of a broad range of areas of mathematics, including dynamics, diophantine approximation, ergodic theory, geometry, and Lie group theory.

Marina was born in Moscow on October 30, 1938 to scientist parents, her father a plant physiologist and mother a chemist. As a Jewish family they had a difficult time in Russia at that time. In particular, her own mother lost her job for having corresponded with none other than her mother who was in Israel, which was considered an enemy state. Marina was educated in Moscow and fell in love with mathematics when she was in the fifth grade; "mathematics came naturally to me and I felt unmatched satisfaction solving difficult problems" she was to aver later.[1] After completing school she gained admission to the Moscow State University, which, with the dawning of the Khrushchev era, had begun to accept Jewish students on an equal footing.

After graduating from the University in 1961, Ratner worked for four years as an assistant in the Applied Statistics Group of A. N. Kolmogorov, the celebrated Russian mathematician who laid the foundations of measure-theoretic probability theory and had a great influence on her during her undergraduate years. Kolmogorov had an intensive training program for talented high school students with which Marina was actively involved. It is also during these years that she gave birth to a daughter from a short-lived marriage.

[1] *M. Cook, Marina Ratner, in:* Mathematicians: An Outer View of the Inner World, *Princeton University Press, 2009, pp. 90–91.*

In 1965 Marina took up research under the supervision of Ya. G. Sinai, a former student of Kolmogorov, renowned, in particular, for his role jointly with his advisor in the development of the very influential "entropy" invariant in ergodic theory around 1960.[2] In the context of how the theory was developing then, in Russia, the geodesic flows associated with surfaces of negative curvature had emerged as crucial examples for study from an ergodic-theoretic point of view, and Ratner also wrote her thesis on this topic.[3] Apart from the examples themselves, a general class of systems known as Anosov flows, named after D. V. Anosov who introduced and proved some deep results about them, were of interest, and Ratner worked on the asymptotic statistical properties of these flows as well. For the work she received the equivalent of the PhD degree in 1969 from Moscow State University.

After receiving the degree Ratner was employed as an assistant at the High Technical Engineering School in Moscow. In 1970 the government of USSR was led, in the face of international pressure, to increase substantially the emigration quotas, sparking an exodus of Russian Jews to Israel, of which earlier there had just been a trickle. Notwithstanding the relaxation in the policy, the government and the bureaucracy in general were highly resistant to emigration and treated those desirous of migrating with utmost severity in various ways. Thus, when Ratner applied for a visa that year (1970) to emigrate to Israel, she was dismissed from the job at the Engineering School.

Marina Ratner giving a talk at the International Colloquium at TIFR, 1996.

Ratner landed in Israel, with her daughter, in 1971 and served as a lecturer until 1974 at the Hebrew University of Jerusalem and then at the Pre-academic School of the Hebrew University for another year. During this period Ratner continued her work on the geodesic flows, and also their generalizations in higher dimensions, and established in particular a property manifesting complete randomness of behavior of the trajectories of the Anosov flows, known as the Bernoulli property, that was much sought after in various systems.

In the West the study of flows analogous to Anosov flows, called "Axiom A" flows, was introduced by Stephen Smale, then at the University of California (UC), Berkeley. Apart from a certain generality of setting, this study involves separating the role of the measure and understanding the dynamics in terms of the construction of special kinds of partitions of the space, known as Markov partitions. Profound work was done in this direction by one of Smales's students, Robert E. Bowen, known commonly by his adopted name "Rufus" Bowen; the work led Bowen to the construction of invariant measures inherently associated to the systems in more general settings, now known as Bowen measures. Bowen completed his doctorate in 1970 and joined the Berkeley faculty in the same year. Not surprisingly, Bowen was interested in the work of Marina Ratner, and their correspondence during her Jerusalem years culminated in Ratner getting an invitation from UC Berkeley, which she joined in 1975 as acting assistant professor.

Another class of flows, called horocycle flows, are seen to have become a major love for Marina after moving to Berkeley. The geodesic flow associated with a surface of constant negative curvature has two natural companion flows, called the contracting horocycle flow and the expanding horocycle flow; they are actually twins, interchangeable through time reversal of directions at each point, so one may simply talk of *the* horocycle flow. Passing through each point of the phase space (consisting of a point of the surface together

[2] *A dynamical system in the present context means a one-parameter group of transformations of a space (sometimes called the phase space), with the parameter representing time (which could be continuous or discrete); the theory focuses on the long-term behavior of the trajectories of points under application of the transformations, namely, as the time parameter tends to infinity. In ergodic theory the phase space is further considered equipped with a measure of unit mass, namely, a probability is associated for points to belong to various subsets; the system is said to be measure-preserving if the probability remains unchanged when any of the transformations under consideration is applied to a point. In these instances one often focuses on trajectories of "generic points" in terms of the measure or equivalently in statistical terms with respect to the initial point. The Kolmogorov–Sinai entropy is a nonnegative number associated with each measure-preserving system, and when the entropy of two systems is different, their long-term behavior is different. The invariant thus enabled distinguishing dynamical systems on a much finer scale than was possible before.*

[3] *The "geodesic flow" consists of starting with a given point on the surface (or manifold in general) and a direction from that point and moving on, along the distance-minimizing paths corresponding to the geometry of the surface, for the desired amount of time, and noting the point of arrival and the direction of movement at that point; thus, the phase space in this case is formed of pairs consisting of a point of the surface and a direction at that point, and the above procedure describes how the transformation is defined. On the usual sphere the trajectories (paths) of such a flow would follow the great circle, and return to the original point after a fixed amount of time. However, when the surface has negative curvature, the trajectories move away from each other substantially, exponentially in time, and when the surface is compact, "almost all" of them (statistically) tend to fill up the whole space with the passage of more and more time.*

with a direction at the point), there is a uniquely defined curve such that if we pick two points on any one of these curves and consider their trajectories under the geodesic flow we find them getting closer and closer with the passage of time, with the distance between the corresponding points of the trajectories tending to zero. Moreover, there is a natural parametrization on these curves with respect to which they can be thought of as the trajectories of a measure-preserving flow, and that is the (contracting) horocycle flow associated with the surface; the expanding horocycle flow arises similarly from consideration of trajectories of the geodesic flow in the reverse direction. These flows have historically proved to be very useful in studying the properties of the geodesic flows. While in the nature of things the horocycle flow would seem just a sidekick of the geometrically majestic geodesic flow, in the theory of dynamical systems the former has acquired a stature of its own, on account of some of its unique properties.


Marina Ratner with M. S. Raghunathan and S. G. Dani at the conference held in her honor at the Hebrew University of Jerusalem in October 2013.

In two papers published in 1978 and 1979 Ratner showed that the horocycle flows are "loosely Bernoulli" while their Cartesian squares are not "loosely Bernoulli"; the loose Bernoullicity property was introduced by J. Feldman, a colleague at Berkeley, and concerns the flow being similar to the standard winding line flows on the torus along lines with irrational slopes, if one allows the time parameter associated with the trajectories to be modified suitably. The fact, as established by Ratner, that the Cartesian square is not loosely Bernoulli for the horocycle flows is rather curious and was the first such instance to be found.

The early 1980s saw a major breakthrough in the understanding of the horocycle flows associated with compact surfaces, of constant negative curvature, at the hands of Ratner. A major question involved was the following: given two such surfaces whether the horocycle flows associated with them being isomorphic to each other as measure-preserving flows would imply that the surfaces themselves are geometrically indistinguishable. The answer to the corresponding question in the case of the geodesic flows is a definitive no since in fact the flows being Bernoullian means (by a well-known result of D. S. Ornstein) that any two of them are isomorphic to each other (up to a rescaling of the time parameter), irrespective of the specific geometries of the underlying surfaces. Ratner proved that on the other hand the horocycle flows corresponding to two distinct compact surfaces of constant negative curvature would never be isomorphic. This kind of phenomenon is referred to as rigidity. She also exhibited various variations of the rigidity property of the horocycle flows, through a series of papers, describing their factors, joinings, etc. (joining is a technical construction that enables comparing two systems with regard to the nature of their dynamics). Two of the three papers in this respect appeared, in 1982 and 1983, in the *Annals of Mathematics*. Apart from the immediate outcomes, which were striking in themselves, the work has germs of the ideas involved in the later celebrated work on the Raghunathan conjecture.

Let me now come to the Raghunathan conjecture, resolution of which was the major feat of Marina's work. Genesis of the conjecture is intricately connected with my student years at the Tata Institute of Fundamental Research, and it would be worthwhile to recall some details in that regard. I did my doctoral work in the early 1970s under the supervision of M. S. Raghunathan on flows on homogeneous spaces.[4] The thesis dealt primarily with the Kolmogorov property, which is a statistical property concerning a strong form of mixing, with no direct bearing on the behavior of individual orbits. However, in a paper written shortly after completing the thesis paper (before the award of the degree, in fact) I proved that all the orbits of actions of a class of flows, more specifically horospherical flows, are dense in the space. Around that time Jyotsna Dani (my wife) who was working under the supervision of S. Raghavan, at TIFR, had proved that for any vector whose coordinates are nonzero and not rational multiples of each other, the

[4]*For an idea of homogeneous spaces and dynamics on them let us consider a Euclidean space and agree to identify two given vectors of the space if their difference has integer coordinates, namely, we view the vectors modulo the lattice of vectors with integral coordinates; geometrically in effect we are considering a torus, and translations by vectors on any particular line define a translation flow on the torus. Similarly, when the elements of various matrix groups, more generally Lie groups, are considered modulo elements of large enough discrete subgroups (called lattices) we get what are called homogeneous spaces with a natural finite measure on them, and matrix multiplication by elements from a one-parameter subgroup of the ambient group, considered modulo the lattice, defines a flow on the homogeneous space. In the particular case when the group involved is the "modular group," the group of 2×2 matrices with real entries and determinant 1, the flows arising in this way (other than those which are periodic) in fact correspond to the geodesic and horocycle flows associated with various surfaces with constant negative curvature, via certain natural identification of the phase space of the flow with a homogeneous space of the modular group.*

Marina Ratner at the inaugural function of the International Colloquium on Lie Groups and Ergodic Theory at the Tata Institute of Fundamental Research, Mumbai, 1996. The others on the dais, from left to right are, S. G. Dani, Virendra Singh, R. Chidambaram, Hillel Furstenberg, Anatole Katok, and M. S. Raghunathan.

orbit under the action of the group of integral unimodular (determinant 1) matrices on the corresponding Euclidean space, is dense in the Euclidean space. At some point in time around 1975, which had these events in the background, when I was talking to Raghunathan about possible problems to pursue, he casually suggested a statement on the behavior of what are called unipotent flows[5] and quite nonchalantly added "call it my conjecture and prove it." He pointed out that proving it would in particular settle the conjecture of Oppenheim on density of values of indefinite forms at integral points,[6] which was one of the hallowed problems at that time in the Tata Institute precincts.

That statement of Raghunathan—the Raghunathan conjecture—first recorded in print in my *Inventiones Mathematicae* (1981) paper, is that the closure of any orbit of a unipotent one-parameter subgroup acting on a homogeneous space of finite volume is the orbit of a (possibly larger) subgroup of the ambient Lie group; in particular this means that each of these closures of orbits is a geometrically nice object—this is a remarkable thing to happen for a dynamical system, the crucial point being that the statement is being made for *every* orbit and not only the generic ones.[7] In that paper I proposed another conjecture, as a step toward proving the Raghunathan conjecture, relating to measures that are invariant under these flows, namely, that the ergodic[8] ones from among them in fact arise as measures invariant under the action the larger subgroups as above and are supported on a single orbit of the subgroup; a weak result was proved in the paper in that direction, partially vindicating the conjecture. Ratner proved the latter conjecture, which she referred to as "Raghunathan's measure conjecture," and in a separate paper deduced the original topological version.[9]

The Oppenheim conjecture itself, which had inspired the Raghunathan conjecture, was settled by G. A. Margulis in 1986 by proving a much weaker statement than the Raghunathan conjecture but in a similar spirit. Not surprisingly, proving the full conjecture led to a much broader perspective in the study of values of quadratic forms at points with integer coordinates, and many other applications, some quite immediately and many more over the years. There have been numerous results since then making use of Ratner's theorems in crucial ways, in a variety of contexts, and there is no doubt that it will serve as a mainstay for a good deal of mathematics in the coming decades.

The proofs are long and intricate and involve various ancilliary results. However, there is a beautiful key idea that concerns observing and adopting a property of the unipotent flows, which it may be worthwhile to recall. It may be informally stated as the following: if you find two trajectories of the flow having stayed quite close for reasonably long, then you can expect them to stay fairly close for substantially longer. This property of the unipotent flows, now called the Ratner property, has since acquired significance as a dynamical phenomenon.

As to be expected, Ratner gained considerable professional recognition. While her initial appointment at Berkeley had been a source of some controversy in the Department, her subsequent rise in the ranks seems to have

[5]*A flow on a homogeneous space of a matrix group as in the previous footnote is said to be unipotent when the one-parameter group involved consists of* unipotent *matrices, namely, matrices that have no eigenvalue, even in complex numbers, other than 1; for a general Lie group there is a variation of this involved. In the case of the modular group there are precisely the horocycle flows associated with surfaces of constant negative curvature and finite area.*

[6]*The conjecture originating from a paper of Alexander Oppenheim from 1929 predicted that for any nondegenerate indefinite quadratic form in at least three variables, which is not a multiple of a form with rational coefficients, the set of its values at integer tuples is dense in real numbers. It had been worked on by several notable number theorists, and by the 1980s many partial results were known, confirming the conjecture under various restrictions, but a general solution had eluded the efforts.*

[7]*For a one-parameter flow the orbit of a point is the set of all the points that can be reached by application of one of the transformations from the flow (including those corresponding to the negative value of the time parameter); similar terminology applies also to a more general group of transformations, in place of the one-parameter flows. The closure of the orbit means all the points that can be approximated by points on the orbit. In a typical dynamical system, even when the closures of almost all orbits are the whole space, for others, the exceptional ones, the closures can be very crazy. For instance, for the geodesic flows as in the above discussion there are orbit closures whose intersection with some curves transversal to the flow consists of a mess of uncountably many individual points disconnected from each other.*

[8]*Those that cannot be expressed nontrivially as a sum of two invariant measures.*

[9]*During the interim there were various partial results proved in that direction, but we shall not concern ourselves with it here.*

been smooth-sailing. She was elected in 1992 to the American Academy of Arts and Sciences, and in 1993 she was awarded the Ostrowski Prize.[10] In 1994 she won the John J. Carty Prize of the National Academy of Science. She was invited as a plenary speaker at the International Congress of Mathematicians, held in Zurich, in 1994, to become only the third woman mathematician, along with Ingrid Daubechies, to receive such an honor; Emmy Noether (in 1932) and Karen Uhlenbeck (in 1990) are the two women to have received the distinction earlier.

Marina Ratner at an excursion to Lake Louise, Alberta, Canada, with some of the delegates to the Conference on Ergodic Theory, held at the Banff International Research Station, Banff, Alberta, Canada, in July 2005. In the photo, from left to right, are Dave Witte Morris, Nimish Shah, Marina Ratner, S. G. Dani, and M. S. Raghunathan.

A conference on "Homogeneous Dynamics, Unipotent Flows, and Applications" was held at the Hebrew University of Jerusalem, October 13–17, 2013, in honor of Marina Ratner and her work, hosted by the Israel Institute for Advanced Studies and supported by the European Research Council. Earlier that year the Hebrew University of Jerusalem conferred upon her an honorary doctorate, at its Convocation held on June 16, 2013.

My personal contacts with Marina were, unfortunately, only sporadic, though they extended over a stretch of more than three decades.[11] I found her a very warm-hearted person, going out of her way to extend hospitality, which I had numerous occasions of enjoying together with my family. We had an International Colloquium on Lie Groups and Ergodic Theory at TIFR in 1996 and the pleasure of having Ratner as one of the speakers. She also contributed a paper to the proceedings of the colloquium on p-adic and S-arithmetic generalizations of the Raghunathan conjecture.

My last meeting with her was in 2015, when there was a special semester organized at the Mathematical Sciences Research Institute, Berkeley, on homogeneous dynamics. On one evening she had organized a dessert party at her home. It had been a wonderful evening thanks, apart from the sumptuous desserts of wide variety, to the warm reception by Marina that she conducted so cheerfully and energetically. My wife and I got to see some photographs from her visit to Mumbai, which she had dug out for the occasion. We also got to meet her daughter Anna and her children. On receiving the news of her sad demise, I emailed Anna a condolence message expressing shock and sadness, in which I also mentioned how energetic Marina had seemed at the party. In her response Anna added, "This is all very sudden and unexpected and difficult to comprehend. She was always so full of energy." Indeed, her sad demise was very abrupt, and we deeply miss her lively presence amongst us.

Credits

Opening photo of Marina Ratner is courtesy of Anna Ratner.

Photos of Marina Ratner at the 1996 International Colloquium at TIFR are courtesy of TIFR Archives.

Photo of Marina Ratner at the Hebrew University of Jerusalem, 2013 is courtesy of Israel Institute for Advanced Studies, The Hebrew University of Jerusalem.

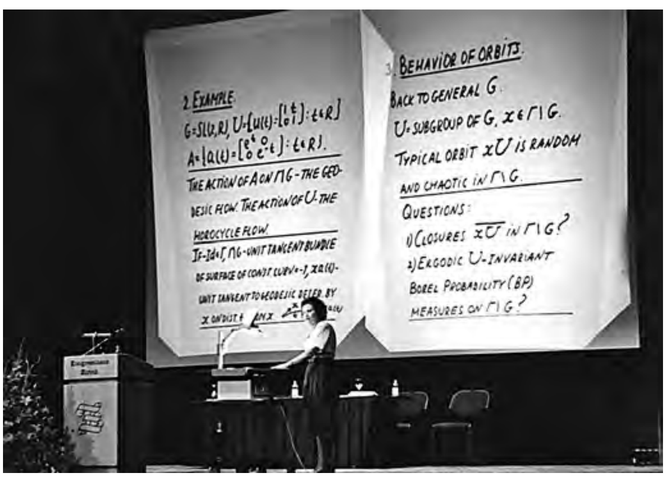
Photo of Marina Ratner at Lake Louise is courtesy of Nimish A. Shah.

[10] *The prize is awarded, since 1989, by the Ostrowski Foundation every alternate year for outstanding achievements in pure mathematics or foundations of numerical analysis.*

[11] *The first of these was in the spring of 1982 when I had an opportunity to visit the University of California, Berkeley for the semester; though in anticipation of the visit I was hoping for a serious mathematical interaction with her, it turned out, much to my disappointment, that she was on sabbatical leave during the period, which she was spending at Stanford University, and we happened to meet only occasionally during her brief visits to Berkeley.*

Ratner's Work on Unipotent Flows and Its Impact

Elon Lindenstrauss, Peter Sarnak, and Amie Wilkinson



Ratner presenting her rigidity theorems in a plenary address to the 1994 ICM, Zurich.

In this note we delve a bit more into Ratner's rigidity theorems for unipotent flows and highlight some of their striking applications, expanding on the outline presented by Dani above. As the name suggests, these theorems assert that the closures, as well as related features, of the orbits of such flows are very restricted (rigid). As such they provide a fundamental and powerful tool for problems connected with these flows. The brilliant techniques that Ratner introduced and developed in establishing this rigidity have been the blueprint for similar rigidity theorems that have been proved more recently in other contexts.

We begin by describing the setup for the group of $d \times d$ matrices with real entries and determinant equal to 1 — that is, $\mathrm{SL}(d, \mathbb{R})$. An element $g \in \mathrm{SL}(d, \mathbb{R})$ is *unipotent* if $g - 1$ is a nilpotent matrix (we use 1 to denote the identity element in G), and we will say a group $U < G$ is unipotent if every element of U is unipotent. Connected unipotent subgroups of $\mathrm{SL}(d, \mathbb{R})$, in particular one-parameter unipotent subgroups, are basic objects in Ratner's work. A unipotent group is said to be a *one-parameter unipotent group* if there is a surjective homomorphism defined by polynomials from the additive group of real numbers onto the group; for instance

$$u(t) = \begin{pmatrix} 1 & t \\ & 1 \end{pmatrix} \qquad \text{and} \qquad u(t) = \begin{pmatrix} 1 & t & t^2/2 \\ & 1 & t \\ & & 1 \end{pmatrix}.$$

In both cases it is easy to verify directly that these polynomials do indeed define a homomorphism: i.e., for any $s, t \in \mathbb{R}$ it holds that $u(t+s) = u(t) \cdot u(s)$. While there is essentially no loss of generality in discussing only the case of $\mathrm{SL}(d, \mathbb{R})$, a more natural context is that of linear algebraic groups — subvarieties of $\mathrm{SL}(d, \mathbb{R})$ defined

Elon Lindenstrauss is Alice Kusiel and Kurt Vorreuter professor of mathematics at The Hebrew University of Jerusalem. His email address is elon@math.huji.ac.il.

Peter Sarnak is Eugene Higgins professor of mathematics at Princeton University and also a professor at the Institute for Advanced Study. His email address is sarnak@math.princeton.edu.

Amie Wilkinson is a professor of mathematics at the University of Chicago. Her email address is wilkinso@math.uchicago.edu.

For permission to reprint this article, please contact: reprint-permission@ams.org.

DOI: https://doi.org/10.1090/noti/1829

by polynomial equations that are closed under multiplications and taking inverses (this notion actually makes sense for more general fields than the real numbers; if we want to emphasize that we are working with the field of real numbers we will call such groups linear algebraic groups *over* $\mathbb{R}$). Connected unipotent subgroups of $\mathrm{SL}(d, \mathbb{R})$ are always linear algebraic groups. Another nice class of examples are the orthogonal groups. Given a quadratic form $Q(\mathbf{x})$ over $\mathbb{R}$ (positive definite or not) in d variables, one can consider the group $\mathrm{SO}(Q)$ of all matrices in $\mathrm{SL}(d, \mathbb{R})$ that preserve this form, i.e. $d \times d$-matrices M so that $Q(M\mathbf{x}) = Q(\mathbf{x})$ for all $x \in \mathbb{R}^d$. This group will be compact if and only if Q is a positive definite or a negative definite form.

Ratner's theorems on rigidity of unipotent group actions deal with the action of a unipotent group U on a quotient space of G by a discrete subgroup. An important example of such a quotient space is when $G = \mathrm{SL}(d, \mathbb{R})$ and $\Gamma = \mathrm{SL}(d, \mathbb{Z})$, in which case G/Γ can be identified with the space lattices in $\mathbb{R}^d$ that have unit covolume. A lattice in $\mathbb{R}^d$ can be specified by giving d linearly independent vectors that generate it — i.e. vectors $v_1, \ldots, v_d \in \mathbb{R}^d$ (that we prefer to think of as column vectors) so that $\Lambda = \mathbb{Z}v_1 + \cdots + \mathbb{Z}v_d$, and the condition that the lattice has unit covolume amounts to requiring that $\det(v_1, \ldots, v_d) = 1$, or in other words that the matrix $g = (v_1, \ldots, v_d)$ obtained by joining together these d vectors be in $\mathrm{SL}(d, \mathbb{R})$. The generators of the lattice Λ are not uniquely determined: $v'_1, \ldots, v'_d$ generate the same lattice as $v_1, \ldots, v_d$ if and only if $(v'_1, \ldots, v'_d) = (v_1, \ldots, v_d)\gamma$ for $\gamma \in \mathrm{SL}(d, \mathbb{Z})$, in other words, lattices of unit covolume in $\mathbb{R}^d$ are in one-to-one correspondence with elements of $\mathrm{SL}(d, \mathbb{R}) / \mathrm{SL}(d, \mathbb{Z})$. Any matrix $h \in \mathrm{SL}(d, \mathbb{R})$ acts on this space by left multiplication; in terms of lattices this amounts to the map from the space of unit covolume lattices to itself taking a lattice $\Lambda < \mathbb{R}^d$ to the lattice $\{h.v : v \in \Lambda\}$.

This quotient space has the important property of having finite volume, or more precisely an $\mathrm{SL}(d, \mathbb{R})$-invariant *probability* measure. A subgroup Γ of a topological group G which is discrete and such that G/Γ has finite volume is called a lattice (admittedly, this can be a bit confusing at first since our basic example of such G/Γ is the space of lattices in $\mathbb{R}^d$..., though this terminology is consistent). Hermann Minkowski seems to have been the first to realize the importance of such quotients, and in particular the space of lattices in $\mathbb{R}^d$, to number theory at the turn of the 19th century. In the introduction to his book *Geometrie der Zahlen*, Minkowski writes[1]

> *This book contains a new kind of applications of analysis of the infinite to the theory of numbers or, better, creates a new bond between these two areas... Geometry of Numbers is how I have called this book, since I arrived at the methods, which deliver in it proofs of arithmetic theorems, through spatial considerations.*

Ratner's work is a remarkable contribution in the general theme of applying "analysis of the infinite" and "spatial considerations" to number theory.

So what did Ratner prove in these remarkable papers? Perhaps the easiest to explain is her Orbit Closure Classification Theorem, confirming an important conjecture of M. S. Raghunathan:

Theorem 1 (Ratner's Orbit Closure Theorem [M3]). *Let G be a real linear algebraic group as above, Γ a lattice in G and $U < G$ a connected unipotent group. Then for any point $x \in G/\Gamma$ the closure of its U-orbit is a very nice object: a single orbit of some closed connected group L that is sandwiched between U and G (and may coincide with either). Moreover, this single orbit of L has finite volume.*

Recall that the U-orbit of a point x is simply the set $\{u.x : u \in U\}$. Note that in particular this shows that any U-orbit closure has a natural U-invariant probability measure attached to it. We also remark that one can loosen the requirement that U be unipotent to U being generated by one-parameter unipotent groups — the passage from Theorem 1 to this more general statement is not very difficult. Unlike previous work towards Raghunathan's Conjecture, in particular Margulis' proof in the mid 1980s of the (then) fifty year old Oppenheim Conjecture using a special case of Raghunathan's Conjecture, Ratner's route to classifying orbit closures was not direct but by via a measure classification result:

Theorem 2 (Ratner's Measure Classification Theorem [M2, M1]). *Let G, Γ and U be as in Theorem 1. Then the only (Borel) probability measures on G/Γ that are invariant and ergodic under U are the natural measures on the orbit closures described in Theorem 1.*

This requires a bit of explanation: We equip $X = G/\Gamma$ with the Borel σ-algebra $\mathcal{B}$, and consider probability measures on the measurable space $(X, \mathcal{B})$. Such a measure μ is U-invariant if the push forward of it under left multiplication by every $u \in U$ remains the same; μ is U-ergodic if every U-invariant Borel subset of X is either null or conull. Every U-invariant probability measure can be presented as an average of ergodic ones, hence classifying the U-ergodic measures gives a description of all U-invariant probability measures on X. Dani conjectured this measure classification result in the same paper where Raghunathan's Conjecture first appeared.

It is possible to reduce both Theorem 1 and Theorem 2 to the case where U is a one-parameter unipotent group. The following theorem implies both of the theorems quoted

[1] *Translated from the original German to English.*

above in the one-parameter case, but is used by Ratner as a bridge allowing her to pass from the measure classification theorem (which, as we said in the outset, is the heart of her work on unipotent flows) to the obit closure theorem:

Theorem 3 (Ratner's Distribution Rigidity Theorem [M3]). *Let G, Γ, U be as above, and let $x \in G/\Gamma$. Then there is a U-ergodic probability measure m_x of the form given above (i.e. the uniform measure on a finite volume orbit of a connected group sandwiched between U and G) so that x is in the support of m_x and for any bounded continuous function f on G/Γ we have that the ergodic averages*

$$\frac{1}{T}\int_0^T f(u(t).x)\,dt \to \int f\,dm_x \qquad \text{as } T \to \infty. \tag{0.1}$$

The reader with some basic knowledge of ergodic theory might be fooled to think that (0.1) is an application of the Birkhoff Pointwise Ergodic Theorem. Not so! The Birkhoff Pointwise Ergodic Theorem only gives information about *almost every* point (with respect to a given ergodic measure). The whole point of Ratner's Distribution Rigidity Theorem is that it is true for each and every $x \in G/\Gamma$. Almost everywhere results are almost always much easier to prove,[2] but in a mathematical manifestation of Murphy's Law, such results might say something about virtually all points but if you are given a specific point and want to study its behaviour under a given action they tell you absolutely nothing. To give a simple analogy, it is trivial to prove that for a.e. $x \in [0,1]$ the asymptotic density of occurrence of each of the digits 0,1,2,...,9 in the decimal expansion of x is 1/10, but asking whether this holds for particular numbers of interest such as $2^{1/3}$ or π seems at present to be a hopelessly difficult question!

As it turns out, for some of the most juicy applications of these rigidity results a more general setup is required. To begin with, one may consider linear algebraic groups over other fields; and since the topological structure is very much in play here, the natural class of fields to look at are local fields, i.e. topological fields whose topology is locally compact, such as $\mathbb{R}$ or the p-adic numbers $\mathbb{Q}_p$. Both Ratner [M4] and independently Margulis and Tomanov [GAGM] extended the above results to this setting, and more generally to quotients G/Γ where $G = \prod_{i=1}^k G_i$ with each G_i a linear algebraic group over a local field of characteristic zero.[3] We shall refer to such quotient spaces G/Γ as *S-arithmetic* quotients, a terminology that probably needs some explanation which we omit to avoid too much of a digression. It would have been interesting to have such rigidity results also for local fields of positive characteristic such as $\mathbb{F}_q((t))$ — the field of formal Laurent series with coefficients in the finite field $\mathbb{F}_q$ with q elements — but there seem to be serious technical obstacles to doing so and only partial results in this direction are known.

The rigidity theorems of Ratner have had numerous applications in many areas of mathematics. A highly nontrivial special case of her general measure classification result, namely the classification of measures on a reducible product $(\mathrm{SL}(2,\mathbb{R})/\Gamma_1) \times (\mathrm{SL}(2,\mathbb{R})/\Gamma_1)$ invariant under a one-parameter unipotent group (the interesting case is classifying measures that project to the uniform measure on each $(\mathrm{SL}(2,\mathbb{R})/\Gamma_i)$ factor, or in the ergodic theoretic terminology, *joinings*) was proved by Ratner already in the early 1980s. The original motivation of Ratner in studying these flows was to understand better (and give natural examples for) a property of measure preserving systems called Loosely Bernoulli — we can view this somewhat anachronistically as an application of unipotent flows to the abstract theory of dynamical systems. Since then her work has had several other applications to abstract ergodic theory and descriptive set theory. There are very striking applications of her work to mathematical physics, for instance in the work of Marklof and Strömbergsson on the Lorentz gas, and to geometry. In this note we have chosen to highlight a couple of the many applications of her theorems (as well as the extension to products of linear groups over local fields as above) to number theory.

In making his famous conjecture, Raghunathan was motivated by the connection to the Oppenheim Conjecture, a connection that allowed Margulis to resolve this longstanding open problem by establishing a special case of the conjecture posed by Raghunathan [G]. Oppenheim conjectured in the 1930s that for any indefinite quadratic form Q in $d \geq 3$ variables that is not proportional to a quadratic form with integer coefficients, the set of values attained by Q at integer vectors, that is to say $Q(\mathbb{Z}^d)$, contains zero as a *non-isolated* point. Using Ratner's Measure Classification Theorem, and relying upon prior work by Dani and Margulis, Eskin, Margulis, and Mozes [AGS] were able not only to show that there are integer vectors $\mathbf{n} \in \mathbb{Z}^d$ for which $Q(\mathbf{n})$ is close to a given value (say 0), but to *count* the number of such vectors. More precisely, for indefinite quadratic forms as above, not of signature (1,2) or (2,2), Eskin, Margulis, and Mozes show that for any $a < b$, the number of integer vectors $\mathbf{n} \in \mathbb{Z}^d$ inside a ball of radius R for which $a < Q(\mathbf{n}) < b$ is asymptotically given by the volume of the corresponding shape cut by the two hypersurfaces $Q(\mathbf{x}) = a$ and $Q(\mathbf{x}) = b$ in this ball. Perhaps an illustration of the delicacy of the question is that this natural statement is false(!) for quadratic forms of signature

[2] *This is a slight pun—"almost everywhere" is used in the above sentence in its precise mathematical sense, whereas "almost always" is used in the ordinary, non-mathematical sense of the phrase...*

[3] *Note that our definitions of unipotent groups and one-parameter unipotent groups make sense over any field, and can be easily extended to the product case, e.g. a subgroup $U < \prod_{i=1}^k G_i$ (with each G_i defined over a different local field) is a one-parameter unipotent group if there is an i so that U is a one-parameter unipotent subgroup of G_i.*

(1,2) or (2,2), though in a follow-up paper Eskin, Mozes, and Margulis were able to prove this estimate for quadratic forms of signature (2,2) under a suitable Diophantine condition, a result which is of interest in the context of the study of the statistics of energy levels of quantization of integrable dynamical systems.

The reason unipotent dynamics is relevant to the Oppenheim Conjecture is that the symmetry group of an indefinite (real) quadratic form with ≥ 3 variables contains (indeed, is generated by) one-parameter unipotent groups. Surprisingly, there is a relatively recent application of the S-arithmetic analogue of Ratner's results to *positive definite, integral* forms.

Legendre's Three Squares Theorem says that a positive integer n can be presented as a sum of three squares if and only if it is not of the form $4^a(8b+7)$,with a, b integers. This is an example of a local-to-global principle: the quadratic form $Q(x,y,z) = x^2+y^2+z^2$ represents an integer n if and only if the congruences $Q(x,y,z) \equiv n \pmod{p^a}$ are solvable for any prime p and any $a \in \mathbb{N}$ (for a given p, consistency of this infinite set of congruences is equivalent to $Q(x,y,z) = n$ being solvable by p-adic integers). In this particular case, only the prime 2 can be an obstacle though there is another restriction on n implicit in the way that we set up the problem — that n is positive — which can be said to come from the "place at infinity," in other words from the necessity that $Q(x,y,z) = n$ be solvable over $\mathbb{R}$.

Legendre's Three Squares Theorem can be viewed as a special case of the following problem: Given a fixed positive definite integral quadratic form Q in many (say k) variables, which quadratic forms Q' in $\ell < k$ variables can be represented by Q? That is to say, when can we find a $k \times \ell$ integer matrix M so that as quadratic forms $Q' = Q \circ M$? For $\ell = 1$ and $Q = x^2+y^2+z^2$ this reduces to the question addressed by Legendre: the form $Q' = nx^2$ can be represented by Q iff n can be written as a sum of three squares. Local solvability — the existence of such matrix M with entries in $\mathbb{Z}_p$ for every p — is an obvious necessary condition that can be verified with a finite calculation.

Hsia, Kitaoka, and Kneser in 1978 established the validity of such a local-to-global principle for representing any form Q' in ℓ variables with sufficiently large square free discriminant by a given form Q in k variables once $k \geq 2\ell + 3$ by using more traditional number theoretic methods. This remained the best result on this very classical problem (essentially dating back to the work of Gauss) until Ellenberg and Venkatesh [JA] were able to use the S-arithmetic extensions to Ratner's Orbit Closure Theorem to very significantly reduce the restriction on k and ℓ to be $k \geq \ell + 5$. While we cannot get into the details of the argument, we note that even if a quadratic form Q is positive definite, hence its symmetry group over $\mathbb{R}$ is compact, over the p-adic numbers in general for $k \geq 3$ variables it would be a non-compact group with plenty of unipotents. In truth, the relevant symmetry group for this case is not the symmetry group of Q but the subgroup of this symmetry group fixing a given quadratic form in ℓ variables, but this is precisely why in this problem one needs to employ p-adics.

An even more surprising application of Ratner's work to number theory was given by Vatsal and Cornut–Vatsal (e.g. [V]). We do not give details here, but in these works families of elliptic curve L-functions, and in particular their central values (or derivatives thereof when their functional equation is odd rather than even), are considered. Using Ratner's Orbit Closure Theorem as a basic ingredient Vatsal (and in the more general cases Cornut and Vatsal) showed that all but finitely many of these values are not zero. When combined with well-known results towards the Birch and Swinnerton–Dyer Conjecture, this proves a conjecture of Mazur: essentially all the points on an elliptic curve whose coordinates lie in ring class fields with restricted ramification are generated by explicit special points first constructed by Heegner.

The impact of Ratner's work cannot be measured only by direct application of her seminal results. Techniques introduced by Ratner to study ergodic theoretic joinings in her early works on unipotent flows in the 1980s were a main inspiration in the work of the first named author on diagonalizable flows and its applications to Arithmetic Quantum Unique Ergodicity and equidistribution. Benoist and Quint were similarly inspired by Ratner's work in their breakthrough work understanding stationary measures and orbit closures for actions of thin groups on homogeneous spaces, and Eskin and Mirzakhani transformed the study of moduli spaces of abelian and quadratic differentials on Riemann surfaces by proving an analogue of Ratner's work in this setting.

The prevalence of deep and suprising applications of Ratner's Rigidity Theorems on unipotent flows is remarkable, and shows the richness of the subject of homogeneous dynamics and how interconnected it is with many other subjects. It is also a tribute to a wonderful mathematician who has left a legacy to future mathematicians for many years to come.

References

[AGS] Eskin A, Margulis G, Mozes S. Upper bounds and asymptotics in a quantitative version of the Oppenheim conjecture, *Ann. of Math. (2)*, no. 1 (147):93–141, 1998, DOI 10.2307/120984. MR1609447

[JA] Ellenberg J, Venkatesh A. Local-global principles for representations of quadratic forms, *Invent. Math.*, no. 2 (171):257–279, 2008, DOI 10.1007/s00222-007-0077-7. MR2367020

[G] Margulis G. Discrete subgroups and ergodic theory, *Number theory, trace formulas and discrete groups* (Oslo, 1987); 1989:377–398. MR993328

[GAGM] Margulis G A, Tomanov G M. Invariant measures for actions of unipotent groups over local fields on homogeneous spaces, *Invent. Math.*, no. 1-3 (116):347–392, 1994, DOI 10.1007/BF01231565. MR1253197

[M1] Ratner M. On measure rigidity of unipotent subgroups of semisimple groups, *Acta Math.*, no. 3-4 (165):229–309, 1990, DOI 10.1007/BF02391906. MR1075042

[M2] Ratner M. On Raghunathan's measure conjecture, *Ann. of Math. (2)*, no. 3 (134):545–607, 1991. MR1135878

[M3] Ratner M. Raghunathan's topological conjecture and distributions of unipotent flows, *Duke Math. J.*, no. 1 (63):235–280, 1991, DOI 10.1215/S0012-7094-91-06311-8. MR1106945

[M4] Ratner M. Raghunathan's conjectures for Cartesian products of real and p-adic Lie groups, *Duke Math. J.*, no. 2 (77):275–382, 1995, DOI 10.1215/S0012-7094-95-07710-2. MR1321062

[V] Vatsal V. Uniform distribution of Heegner points, *Invent. Math.*, no. 1 (148):1–46, 2002, DOI 10.1007/s002220100183. MR1892842

Credits

Photo of Marina Ratner is courtesy of Anna Ratner.

MEMORIAL TRIBUTE

To the Memory of Marina Ratner

Yakov Sinai

Sinai and Ratner in 1978.

I met Marina almost the same time as her elder sister, Yulia. Their father was a famous biologist. Marina started as a student of Moscow State University, and initially she was a student of A. N. Kolmogorov and later became a student of R. L. Dobrushin. When she entered graduate school she became interested in ergodic theory. This is how I became her advisor.

Marina married A. Samoilov when they both were undergraduate students of the second year. Their marriage didn't last long and soon they separated, though they maintained contact. Marina was left with her daughter, Anya. Marina was very close to the family of her daughter. During many years, Marina spent a lot of time with her grandchildren. Marina's friends knew that it was strictly forbidden to call her on Saturdays because she was always busy working the whole day with her grandson.

Yakov Sinai is a professor of mathematics at Princeton University. His email address is sinai@math.princeton.edu.

DOI: http://dx.doi.org/10.1090/noti1830

Marina wrote her thesis about Markov partitions in multi-dimensional systems. At that time it was a very hot topic. One of the referees of Marina's thesis was V. A. Rokhlin, who wrote a very good report. This was important because the Scientific Council where Marina's thesis was considered was known at that time for its antisemitism. It was rather surprising that in the case of Marina the system worked well, including the voting of the PhD committee and the approval of the High Attestation Committee (VAK).

Quite soon she started to work in one of the Moscow Institutes where Marina was appointed to her first academic position. The fact that she got a position so quickly was rather unusual at the time.

In another case it would be a big step in someone's career. But not for Marina, because quite soon she decided to go in a different direction and applied to emigrate to Israel together with her daughter.

In Israel she joined the Institute of Mathematics at the Hebrew University and started teaching there. Marina did everything very well. Soon she became famous among her students. Many of them kept as souvenirs the notes left after Marina's classes.

A bit later Marina heard about some vacancies opening in Berkeley and moved there with her family. Berkeley became her home until the end of her life. She was elected as a full member of the National Academy of Sciences of the USA, and was invited as a plenary speaker to ICM-94 in Zürich.

Marina had many close friends in Berkeley and other places. One can mention Smale, Ornstein, Arnold, Fuchs, Pyatetskii-Shapiro, Kazhdan, Chorin, Zalenko, Kresin, and many others. Marina was always ready to help her friends and other people. I remember the case when our family arrived to Princeton after my son had seriously broken his leg. Marina contacted many people and eventually they helped us to find V. Golyakhowsky, who was a remarkable orthopedist. His treatment was excellent, and my son completely recovered and can freely walk now without any trace of the previous accident.

Marina was involved in many types of social and political activity, and was very strong and principled in promoting the causes that she believed in. For instance, she had strong opinions about mathematical education and education in general.

This text is a small part of what can be written about Marina. She was a great mathematician, a remarkable personality, and a close friend.

Credits

Photo is courtesy of Anna Ratner.

MEMORIAL TRIBUTE

Meetings: For Almost All of Our Lives

Boris Gurevich

Ratner with her daughter Anna, 1971.

Ratner with her family, 1947 (left to right): mother Ksia, sister Yulia, Marina, and father Yehoshua.

I cannot claim that I have been Marina's close friend from our very first meeting. But I believe at some moment this became so. Our meetings lasted for many years, with frequency dependent on circumstances of our lives and on political events as well. I am going to remember several of these meetings in the hope that my story will shed additional light on this remarkable character.

I got to know Marina when we were about seven years old and went to the same musical school for children. By coincidence we had the same piano teacher, whose name was Anna Ratner. In one or two years I moved to another musical school, closer to my home, and we lost one another for several years.

Our next meeting occurred at the Mechanics and Mathematics Department of Moscow State University, where we entered simultaneously and quite independently. I happened to meet Marina at one of the first lectures and recognized her almost immediately, strange though it may seem. That is why when in several days our very sociable

Boris Gurevich is a professor of mathematics at Lomonosov Moscow State University. His email address is gurevich@mech.math.msu.su.

DOI: http://dx.doi.org/10.1090/noti1831

Visiting Leningrad with fellow students—(left to right) Ilya Mindilin, Vener Galin, Lena Odnorobova (or Vera Steniushkina), Slava Perlov)—from Moscow, circa 1961.

fellow student tried to introduce me to her, this sounded funny for both of us.

During the first two years we were in different groups and met only at common lectures. But in our third year, everybody had to choose a specialization. And again, we made the same choice, which was probability, and got into the same group.

At that time the probability and statistics subdivision of the department was headed by A. N. Kolmogorov, and almost all who worked there were his former students. Kolmogorov was very active in various directions; in a few years he included Marina in a small, young team involved in his study of statistical laws in language. But her first supervisor was R. L. Dobrushin, who, as I know, liked very much her master's thesis in information theory.

Upon graduating from the university, Marina was for some time working at Kolmogorov's boarding school, a high school for gifted children from all over the country founded by Kolmogorov and later named after him. She also took part in the preparation of the principal works of Claude Shannon for publication in Russian.

About that time she married a student from our course, and I met her not too often. But in 1965 she came back to the university as a graduate student under the supervision of Y. G. Sinai, and our meetings became regular again because we both attended seminars on ergodic theory.

Once we examined an undergraduate. I began with a question, then Marina entered and I went out for some time, while she continued. When I came back, I was not pleased that she finished too fast. But later I decided that she was right: first, this student was Lenya Bunimovich, and second, Marina was very thorough in everything. Once, already in Berkeley, she observed that in one of her papers, it was written "weak convergence" instead of "weak* convergence," and she asked me to insert the asterisk by pen each time I was at a library where a journal with the paper was accessible. She told me, "I am a perfectionist," and this was the truth.

She published several papers and defended her PhD thesis on geodesic flows in 1969. Thereafter she was teaching at one of the technical universities in Moscow, but not for very long, because she applied for emigration to Israel.

As a girl of twelve, 1950.

Honestly speaking, I initially considered her intention reckless: I knew that she was going alone, with a small daughter, without language and having no relatives there.

But I had underestimated Marina: she overcame difficulties, which were indeed considerable, and in 1971 she was already working at the Hebrew University. At the time, the Soviet Union had no diplomatic relations with Israel and postal services were unreliable. Of fragmentary information from Marina I remember that in the fall of 1973, the university professors were asked to write their lectures down in order that the students, turned into soldiers for a while, could read them at the front.

When Marina moved to Berkeley, I used a possibility to hear something of her from J. Feldman, whom I met in 1977 in Warsaw at a conference on ergodic theory.

Only when Gorbachev came to power did mutual visits become possible. Marina came to Moscow more than one time in the 1990s and early 2000s. Once she left for a few days for Minsk, where a mathematical conference was conducted. Being aware of food shortage in Moscow at the time, she bought in Minsk, on her own initiative, some cheese for a small child of our friend. I appreciated her solicitous concern for her friends once again when I visited her at Berkeley in the late 1990s.

As far as I know, she came to Moscow for the last time in June of 2003 to the conference devoted to Kolmogorov's centennial, where she was an invited speaker and met many old friends.

I saw Marina for the last time in May 2014 in Oslo, where we were invited by Sinai, our common teacher, as his guests as he was awarded the Abel Prize. We walked through the city in full lilac bloom and followed Sinai, visiting the town of Stawanger for one day, where we took, together with a few friends, an excursion along a fjord; Marina took a number of snapshots there.

She was always worried about the health of others. Throughout several years she insisted that I should regularly inform her of my state of health (results of tests, etc.). Answering my questions about her own health, she would always insist that she was splendidly sound. I have kept her last message of July 1, 2017, in which she wrote about her problems, but hoped that the treatment would eventually help. I also thought so.

Credits

All photos are courtesy of Anna Ratner.

In Memory

D. Ornstein

Ratner with (left to right) Ornstein, Yakov Sinai, and Jack Feldman, 1978.

Marina Ratner was a good friend and colleague. Although we never wrote a joint paper, we did work together, and I was able to gain great appreciation for the depth of her mathematical ability. The commitment to her family was very impressive; she homeschooled her grandchildren. She was not a fan of affirmative action and made it very clear that she wanted her achievements to be rated solely on her mathematics not her gender.

I would like to call attention to Marina's work on the horocycle flow, which I hope will not be overlooked in the light of her later and more spectacular results. The study of the horocycle flow and the geodesic flow as flows on an abstract measure space began in 1938 with the work of Hedlund and Hopf. While the geodesic flow is the most random measure preserving flow on an abstract measure space, the horocycle flow has the opposite behavior. Marina elucidated its rigidity properties.

I will give just one example. If two horocycle flows on the quotient space M of $SL(2,\mathbb{R})$ by a discrete subgroup are the same as measure preserving flows, then the underlying surfaces are conformally isometric. This means that even though we replaced M by an abstract measure space, the flow retains all of the geometry that we threw out.

Donald Ornstein is a professor emeritus of mathematics at Stanford University. His email address is `ornstein@math.stanford.edu`.

DOI: http://dx.doi.org/10.1090/noti1832

Credits

Photo is courtesy of Anna Ratner.

MEMORIAL TRIBUTE

Marina Ratner, *quelques évocations*

Jean-Paul Thouvenot

At my first encounter with Marina, in Jerusalem, in 1974, just after the Lavi Conference, I saw a young woman quite shy, sweet and smiling, pleased to receive a gift that a common friend had prepared for her from Paris. I met her later in Berkeley when she was already settled; I paid several visits there. Memory is sometimes strangely selective—I remember distinctly that, in an excursion which we took together with her and the Katok family to "Pebble Beach" (this I am not sure of), she had a very battered car, with the exhaust threatening to fall off at every turn.

Her mathematics, which had started in Moscow with Sinai, received the influence of the California environment, and one of her first works there was the proof that the horocycle flow is loosely Bernoulli, an abstract measure-theoretic property that was quite popular at that time in Berkeley, the impetus for it having been given by Jack Feldman. A second paper, which came quite quickly, was that the Cartesian square of the horocycle flow is not loosely Bernoulli. This was, for the group of people working in this field, quite unexpected and very strong. The non-loosely Bernoulli property all of a sudden being attached to a simple algebraic object, while all previous examples, starting with the one of Jack Feldman, required elaborate combinatorial constructions. This work of Marina is extremely difficult to read, and I remember, when I came to complain (the last time was not so long ago) to her, "But it is so simple, just follow what is written, everything is completely natural, you will not find any obstacle…" In this same work appeared for the first time the "shearing" that was going to play such an important role in her subsequent works. And then came, in an extraordinary succession across a few years of mathematical excitement, a list of papers in which she developed her theory of horocycle flows culminating in the complete description of their joinings, which entails all their rigidity properties. Strikingly, Marina always worked alone and never had coauthors. As to joinings, she managed all by herself, and to my knowledge, without trying to get too many contacts with the people close to her who were active in this field at that time.

On a visit that she paid to Paris at about the same epoch, I put her up in a nice hotel close to Jussieu (as I usually did with visitors). But almost immediately, she came to me quite pleased to have moved with her daughter to a very modest place close to Gare de l'Est, proudly announcing, "Believe me, it is the best possible place as a starting point to visit Paris." Singular Marina!

Her work took a new turning point when she got, as an elaboration of her previous ideas, her fundamental result on the Ragunathan measure conjecture. Strangely, her mathematics, so deep, which is so much alive nowadays, and in so many different directions, is most frequently used as a black box or as a model.

It was a great shock to receive the message from B. Weiss that she had died, shortly before a conference dedicated to the memory of Rufus Bowen, which she had accepted to attend.

I want to mention another memory, or more precisely, an image, because of the deep impression that it has left on me, although I cannot trace back exactly when it took place. I think that it was at a conference in Warwick: she was lecturing, so strong, so determined, and in appearance so fragile, all alone, in front of a huge audience.

In the same way as her mathematics does for our community, her presence, in the minds of those who have known her, persists with all the strength, the singularity, and the seduction of the exceptional.

Jean-Paul Thouvenot is directeur de recherches émérite au CNRS à Sorbonne Université. His email address is `jean-paul.thouvenot@upmc.fr`.

DOI: http://dx.doi.org/10.1090/noti1833

Remembering Marina Ratner

Hee Oh

Ratner with (left to right) François Ledrappier, Dmitry Kleinbock, Hillel Furstenberg, and Oh, 2005.

I first met Marina at the International Conference on Lie Groups and Ergodic Theory held at TIFR in Mumbai in January of 1996. I was a fourth-year graduate student working with Gregory Margulis.

She gave a talk on the p-adic and S-arithmetic generalizations of her earlier proof of Raghunathan's conjecture. I remember how she began her talk with the assertion that while some notations and definitions may be standard, she still needed them to know for herself what she was talking about. She then went on to spend quite a big chunk of her time introducing a long list of notations and basic definitions, such as Ad-unipotents and p-adic Lie algebras. At the time, her talk was too technical for me to follow, but her uncompromising style left a strong impression on me.

My own lecture was about my ongoing thesis work on the arithmeticity of discrete subgroups in a higher rank simple Lie group generated by lattices in a pair of opposite horospherical subgroups, which was a conjecture of Margulis based on Selberg's earlier work in the case of a product of $SL(2,\mathbb{R})$s. I had solved this conjecture for discrete subgroups of $SL(n,\mathbb{R})$ for $n \geq 4$. Ratner's theorem on orbit closures was a key ingredient of my proof. I did not get to receive any comments from her either on my talk or on my work at that time.

Seventeen years later in 2013, there was a conference in her honor titled "Homogenous Dynamics, Unipotent Flows, and Applications" at the Hebrew University. I had just finished my joint work with Amir Mohammadi on the classification of joining measures for geometrically finite subgroups of $SL(2,\mathbb{R})$ or of $SL(2,\mathbb{C})$. It was an extension of her work "Horocycle flows, joining and rigidity of products," published in *Annals of Mathematics* 1983, and our approach was to adapt her proof in the infinite-volume setting. I opened my lecture saying that I was proud of my mathematical aunt; she and Margulis shared a common advisor, Sinai. I then successfully squeezed the two subjects of discrete groups and joinings into my one-hour lecture and closed with the statement that I had started my mathematical career by applying Ratner's theorem as a black box and that I was now hoping to generalize her ideas in the infinite-volume setting. After my lecture, I asked Marina directly, "Did you like my lecture?" She said, "Yes, very much," with a big emphasis on "very," and asked, "Why don't you post your lecture notes in your webpage?" I jokingly replied to her, "Marina, who is going to read it?"

She once said in an email to me, "If a woman is good in math, she does not need encouragement or a role model. I remember when I was young, no matter what anyone would say, I knew that I would go to math. I did not need any encouragement for that. The same is probably true about you. Did you need encouragement?"

I wrote back, saying, "Marina, whether you wanted to or not, you have been a great source of pride and inspiration for female mathematicians in the area. I am very grateful to you for having been such a great role model."

Thank you Marina.

Hee Oh is Abraham Robinson Professor of Mathematics at Yale University. Her email address is `hee.oh@yale.edu`.

For permission to reprint this article, please contact: `reprint-permission@ams.org`.

DOI: http://dx.doi.org/10.1090/noti1834

Credits

Photo is courtesy of Nimish A. Shah.

Creating Opportunities and Building Confidence:
Clare Boothe Luce's Unexpected Support of Women in Math and Science

Della Dumbaugh

ABSTRACT. How did a woman who was a playwright and a politician advance American women in mathematics and science? This paper explores the life of Clare Boothe Luce and her pioneering—and unexpected—impact on the development of mathematics and science.

Introduction

With her death in 1987 Clare Boothe Luce bequeathed nearly $70 million[1] to establish a fund "to encourage women to enter, study, graduate and teach" in the fields of science, engineering, and mathematics. This decision seems an unlikely choice for a woman who, while alive, was widely known as a playwright, magazine editor, American ambassador to Italy, war correspondent, congresswoman, and wife of Henry Luce, who co-founded TIME Inc. Despite having no known connection to or interest in what are now STEM fields [Teltsch], Clare Boothe Luce challenged women to enter into and excel in more commonly male-dominated fields. Her vision established a foundation that has become "the most significant source of private support for women in science, math and engineering in the US [Grant Spotlight]."

The Clare Boothe Luce Program has supported more than 2300 women since awarding the first grants in 1989 [The Clare Boothe Luce Program].[2] The 30th anniversary of the initial Clare Boothe Luce Fund awards provides a timely opportunity to reflect on the life of Clare, to consider her motivation in establishing this support, and to explore the impact of her funding on women and institutions.

Clare Boothe Luce: Life Experiences Shaping a Bequest

On March 10, 1903, in New York City, Clare, born Ann Clare Boothe, began her life as she would live it—surrounded by conflict and drama. Clare was the second illegitimate child of Ann Snyder (Anglicized from Anna Clara Schneider) and William Franklin Boothe [Morris, 1997, p. 15]. William Boothe was legally married to another woman at the time. Although he subsequently divorced his first wife in 1906, William and Ann Snyder never married. After his once successful piano business dwindled, he worked as a medical salesman and, finally, as a musician. In search of work, William's musical career took the family to various cities, including Memphis, Nashville, and Chicago. Money grew increasingly scarce with each move. As William's financial resources faded, so did Ann's affection for him. She had met him as a flourishing executive and now he was an

Della Dumbaugh is a professor of mathematics at the University of Richmond and an associate editor of the Notices. *Her email is* `ddumbaugh@richmond.edu`.

[1] *Roughly $156 million in 2018 dollars.*

DOI: http://dx.doi.org/10.1090/noti1819

[2] *Interestingly, nearly twenty years before her death, Clare proposed the idea of considering a woman for a (Henry) Luce Fellowship. Specifically, in 1968, when asked her opinion on a proposed Luce Fellowship Program at Time, Inc. Clare Boothe Luce wrote mostly about "the man" or "him" in this position. Near the end of the letter, however, she dared to suggest, "Sooo—is there anything in the idea of a Time Inc. Associates Program, among whom, hopefully, the Board of Selection might annually choose a man, or* woman (please!) *worthy to be dubbed a Luce Fellow…" [Clare Boothe Luce to Andrew Heiskell, February 4, 1968, p. 8, Clare Boothe Luce papers, my emphasis].*

aging musician with too few prospects and too much of a drinking habit. Ann Snyder wanted more for her children and for herself. When Ann's father suffered a serious illness in September 1912, she took the opportunity to move her children to her parents' home in New Jersey. She eventually told acquaintances she was a widow. With the death of Ann's father in 1913, the family relocated to New York City [Morris, 1997, p. 39].

This transient lifestyle proved challenging for Clare. She had a difficult time making friends, a situation that would not improve in her lifetime. Clare spent two years at the Cathedral School of St. Mary's in Garden City, Long Island, where some students viewed her as "the most conceited girl in the school [Morris, 1997, p. 57]." Clare felt she would never succeed at St. Mary's, so she appealed to her mother to let her leave. Clare's mother subsequently enrolled her at the Castle School above Tarrytown-on-Hudson in New York. This move was intended to put Clare in a better position to find a suitable husband rather than earn a college degree. At the Castle, although Clare won the school's titles of "Most Artistic," "Cleverest," and "Prettiest," she finished second for "Most Ambitious," the only award she felt she truly deserved. As she expressed it in her diary, "[m]y whole heart and soul is wrapt [sic] up in three things: Mother, Brother and my ambition for success [Clare Boothe Luce Diary, February 6, 1919, as quoted in Morris, 1997, p. 61]."

Clare's drive for success remained with her throughout her life. She decided the best route to success was through marriage, and, in particular, marriage to a wealthy man. As she put it in a letter to a friend, "Damned if I'll ever love any mere man. Money! I need it and the power it brings, and someday you shall hear my name spoken of as—famous [Clare Boothe Luce to Ruth B. Morton, November 18, 1921, as quoted in Morris, 1997, p. 99]."[3] True to her word, Clare loved one man, but married another. At the age of twenty, she married George Tuttle Brokaw, a millionaire alcoholic more than twice her age who simultaneously doubled as New York's most eligible bachelor [Morris, 2014].

Four months after the wedding, Clare learned she was pregnant. Although she tried scalding hot baths as a way to induce an abortion, the child lived and Ann Clare Brokaw was born in 1924. The baby helped the marriage temporarily but could not save a marriage damaged from the start. Clare plotted how to exit the marriage "with minimum damage and the maximum amount of money [Morris, 1997, p. 140]." When Clare and Brokaw amicably divorced in May, 1929,[4] Clare received a settlement of a $425,000 trust fund, an annual income, and expenses for Ann. Following a difficult custody battle, each parent was allotted six months a year with Ann. For all its faults, the marriage provided Clare with plenty of money and an increased social confidence. Although "she had ample means to settle for the life of a socialite" after her divorce from Brokaw, she chose, instead, to "capitalize on her own abilities in the workplace [Morris, 2014, p. 29]."

In an attempt to give her life new direction and meaning, Clare interviewed for a position at *Vogue* magazine. After waiting all summer to hear from the magazine, Clare self-assuredly walked into the *Vogue* building and convinced an office assistant that she was a new employee. Soon enough, colleagues gave the new beautiful, professional woman sitting at an empty desk work to do. *Vogue*'s editor, Edna Woolman Chase, thought the publisher of the magazine, Condé Nast, had hired Clare. Nast, in turn, thought Chase had brought her on board the magazine's staff [Morris, 2014]. Clare received her first paycheck after one month [Morris, 1997, p. 163]. Consequently, with no formal education or experience in writing, Clare secured a job with one of the most popular magazines of the time. She soon moved down the hall to *Vanity Fair* with the title of Junior Editor. Her first piece "Talking Up—and Thinking Down: How to Be a Success in Society Without Saying a Single Word of Much Importance" appeared in 1930 [Clare Boothe Luce, "Talking Up"]. In this article, Clare encouraged readers to be conventional, predictable, safe, and even boring in order to have a successful conversation. She identified the six topics guaranteed to start a conversation: golf, the stock market, prohibition, theater, gossip, and current social activities [Clare Boothe Luce, "Talking Up," p. 39].

After the 1929 stock market crash, *Vanity Fair* struggled to adjust to the new economic conditions. Advertising revenues, for example, dropped twenty percent [Morris, 1997, p. 181]. Clare helped reestablish *Vanity Fair* as a serious magazine concerned with issues beyond the scope of fashion. Her confidence grew with the success of her public-affairs articles. She earned a promotion to associate editor. She used her candor and satire to develop her skills as a political writer. This work led her to the 1932 Democratic National Convention in Chicago, where she met Bernard Baruch, an advisor to Franklin D. Roosevelt and the fourth richest man in America. Baruch introduced Clare to many of the nation's most powerful and prominent men.

With her increasing success, Clare began to take some liberties at *Vanity Fair*. She requested weeks off for personal travel. When in the office, she often arrived late or left early. She produced fewer articles [Morris, 1997, p. 227]. Consequently, Condé Nast expressed concern over her schedule. He also questioned her ability to successfully balance her roles as an editor and author along with her

[3] *As Gore Vidal pointed out more than 75 years later, Clare expressed these thoughts fifteen years before Scarlett O'Hara leapt to and out of the pages of* Gone With the Wind *[Vidal, p. 208].*

[4] *Between 1867 and 1967, the Census Bureau measured the divorce rate by the number of divorces for every 1000 people in the population. In 1929, the rate was 1.7. See* 100 Years of Marriage and Divorce Statistics, 1867–1967.

recent aspirations to become a playwright. These circumstances prompted Clare to leave *Vanity Fair* and begin work as an independent writer. She tried short stories using her trademark satire but found her best work as a playwright. After a few unsuccessful plays, she published *The Women* in 1936 [Luce, *Women*]. *The Women* featured a group of New York's wealthiest idle women whose concerns focused on their physical appearance and the town's latest gossip. Clare worked her progressive views into the play with a conversation between the protagonist and her daughter:

Child: "What fun is there to be a lady? What can a lady do?"

Mother: "These days, ladies do all the things men do. They fly aeroplanes across the ocean, they go into politics and business [Luce, *Women*, p. 23]."

The play opened on Broadway on December 26, 1936 and reached capacity by the end of its fourth week. It ran for 657 performances in the US and 18 countries and grossed over 2 million dollars. The success of *The Women* and two other plays not only established Clare as a talented comedic writer but it also allowed her to embody the life of the modern career woman and encourage others to do the same.

Through her writing, Clare met Henry Robinson Luce, the once humble newspaper reporter on the *Chicago Daily News* now turned publishing magnate with his *Time*, *Fortune*, and *Life* magazines. Harry Luce divorced his wife of 11 years and married Clare in 1935. The marriage lasted 32 years but was not without its challenges. In his *New York Times* obituary, Alvin Krebs suggested that the "rumored difficulties" were "perhaps inevitable in a marriage between two such strongminded personalities [Krebs]."

Although Harry provided Clare with sufficient opportunities to enhance her writing career, Clare now hoped to develop her skills as a politician. In the late 1930s she traveled to Europe to observe political events firsthand. Harry joined her for part of the trip. When she returned to the States, she hastily wrote a nonfiction book titled *Europe in the Spring* to express what she called an eye- and ear-witness report of what she saw [Luce, *Europe in the Spring*]. Her book helped shape public opinion in the US as Americans tried to make sense of the growing crisis in Europe. After the outbreak of war, she accepted the position as War Correspondent for *Life* magazine and traveled again through Europe. These opportunities and her connections allowed Clare to segue into politics.

In 1942, she ran as a Republican in a largely Democratic constituency of the Connecticut district where she lived. She won by a very narrow margin. Women eager to elect the first congresswoman from Connecticut may have earned Clare her victory and she felt honored to fulfill this role. Clare acknowledged that socially established prejudices surrounding women in politics still existed, but she was eager to hold a position with (purportedly) equal opportunities for power and prestige. In Congress,

Figure 1. Congresswoman Clare Boothe Luce of Connecticut.

Clare fought for the continued military strength of the US and she supported equal employment opportunities and racial equality. Of these interests, she prioritized the nation's safety and security above the "feminist issue [Morris, 2014, p. 30]." Clare, however, found it difficult to be taken seriously. While her male colleagues were often valued for their ideas or achievements, she found that female public figures were evaluated on their looks or personalities. As a Congresswoman then, Clare must have found herself at the confluence of the theoretical and the practical, fighting for women's rights while living the reality of a woman in Congress on a daily basis.

Clare felt pressure to succeed. As she put it, "because I am a woman I must make unusual efforts to succeed. If I fail no one will say, 'She doesn't have what it takes.' They will say, 'Women don't have what it takes [Martin, p. 306, Clare's emphasis]."[5] Clare grew tired of politics because she felt politicians were overly critical and never capable of admitting a mistake. She confessed, "I always regretted that I shifted to politics. You can do nothing truly creative

[5] *Clare was inducted into the Connecticut Women's Hall of Fame in 1994 (posthumously). This quote is also featured on her biography page. See* `cwhf.org/inductees/politics-government-law/Clare-boothe-luce#.W5VyKq2ZPUo`.

in politics by yourself [Martin, p. 272]." She continued in her position, however, because she felt she owed it to women to serve as a positive model of an ambitious and successful career woman.

In January 1944, tragedy struck and temporarily put Clare's political frustrations aside. Her daughter, Ann, was killed in a car accident while traveling back to Stanford. Although they had something of a distant relationship, Clare was overcome with grief and regret for not spending more time with Ann. While a student at Stanford, Ann had "pined for her mother to write, telephone or visit. But Clare always had excuses [Morris, 2014, p. 45]." In July 1943, Ann had blamed herself for requesting Clare's attention. "Forgive all my stupid little letters in which I normally ask you to write to me!" Ann wrote to Clare, "[s]omehow I always forget how very busy you are—and all the good you are doing—until I get a batch of clippings! Then it's always a wonder to me how you even manage to survive the work you have to do."[6] When Clare did write to Ann, her affection for her daughter was everywhere apparent. On November 7, 1943, for example, Clare opened her letter with "Annie my pudding-cake, my peach pie, and all assorted delicacies [Morris, 2014, p. 57]." Initially, Clare's grief seemed to propel her into more aggressive and combative types of politics and fueled her 1944 reelection campaign, which she won.

By September 1945, however, Ann's death combined with discouraging world events led Clare to a point of despair [Luce, The Real Reason, April, 1947]. She called (in the middle of the night) a Jesuit priest in New York who had written to her over the years. He referred Clare to Monsignor (later Bishop) Fulton Sheen in Washington, DC. Father Sheen and Clare had several conversations over the course of the next several months and, on February 16, 1946, Clare converted to Catholicism.[7] Since her Connecticut district had a very large Catholic vote, she did not want her newfound faith to be misconstrued as a political maneuver to influence her constituents. To avoid this confusion, two weeks before her conversion, Clare announced that she would not run for Congress again. This decision may have resolved the potential political issue associated with her newly adopted Catholicism, but no matter "how religious Clare became, the loss of Ann remained a persistent and tragic wound [Brenner, p. 164]."

After her two terms as a Congresswoman, she resumed her writing and suffered defeat in a Senate race in 1952. In 1953, Dwight D. Eisenhower appointed Clare ambassador to Italy, the first woman to serve as an American envoy to a major country. The ambassadorship proved mutually beneficial to Clare and to the US. On the diplomatic front, Clare accomplished her three assigned tasks, including advancing the Italian-American friendship, helping to settle the Trieste crisis, and aiding the young democracy of Italy in fighting communism [Hatch, p. 237]. Gore Vidal later went so far as to credit Clare with "single-handedly saving Italy from Communism [Vidal, p. 203]." She retired after this appointment in 1956. She and Harry settled on their ranch in Arizona, although they still traveled extensively. Harry died unexpectedly of a heart attack in 1967. In his will, Harry had established a trust for Clare that paid her interest only, "the absolute minimum he could get away with without having the will challenged [Brenner, p. 166]." The trust would revert to the Henry Luce Foundation at the time of her death. Harry Luce's son, Hank, however, allowed Clare to determine how she would like to use the trust.[8]

The Vision for the Awards: The Clare Boothe Luce Fund

With this freedom, Clare directed the majority of the proceeds of her Estate to support the Clare Boothe Luce Fund "dedicated exclusively to funding scholarships and professorships for women students and professors at educational institutions, a minimum 50% of which shall be Roman Catholic. The purpose of the Clare Boothe Luce Fund shall be to encourage women to enter, study, graduate, and teach in the following fields of endeavor: Physics, Chemistry, Biology, Meteorology, Engineering (Electrical, Mechanical, Aeronautical, Civil, Nuclear and other Engineering disciplines), Computer Science, and Mathematics [CBL Last Will and Testament, p. 12]." Her choice of scientific fields was deliberate. "I select such fields of endeavor in recognition that women today have already entered the fields of medicine, law, business and the arts, and in order to encourage more women to enter the fields of science [CBL Last Will and Testament, p. 12]." The awards were (and are) designated for scholarship and teaching in the US only.[9]

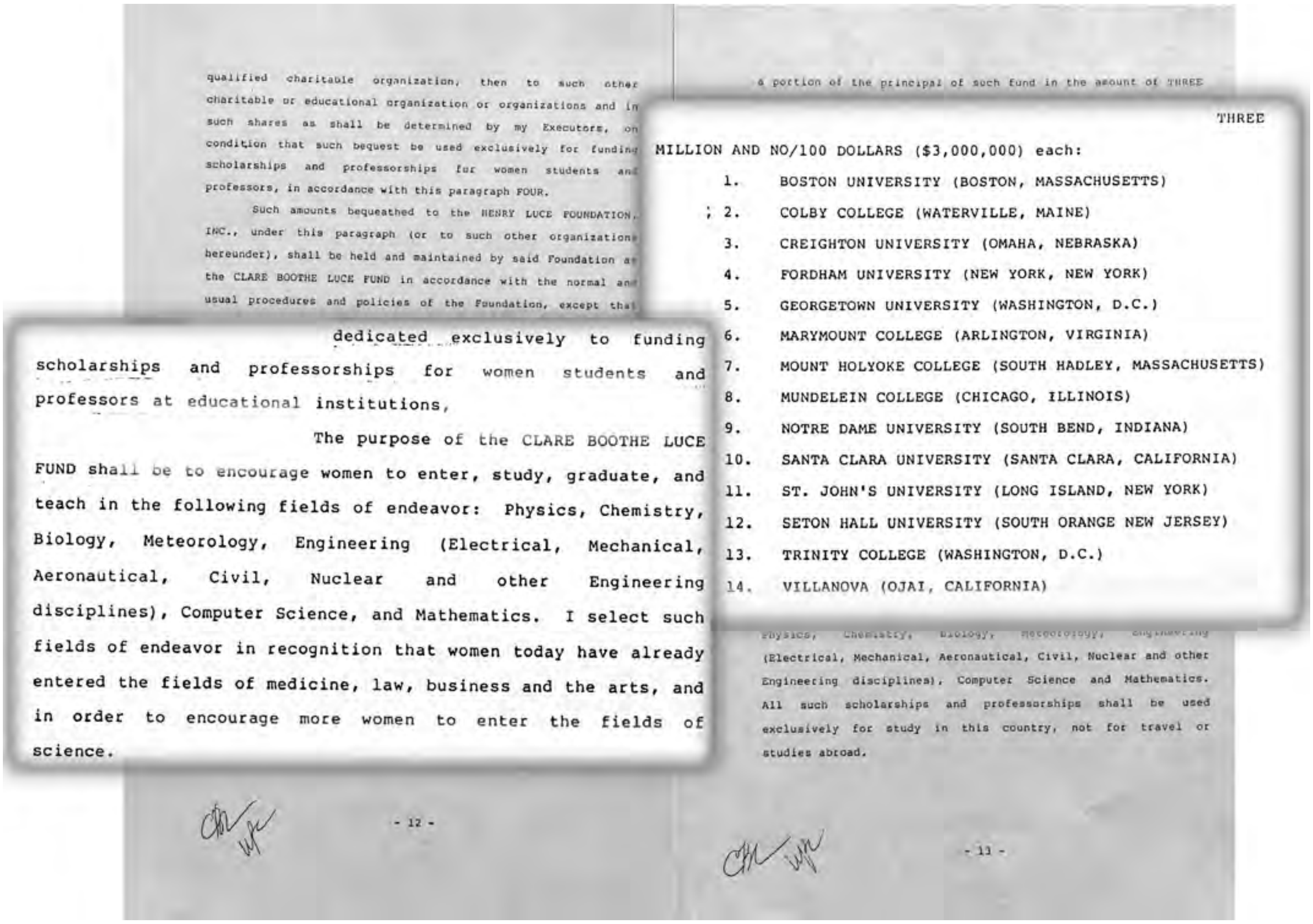
Just as Clare always hoped to accomplish more in life, she hoped other women would do the same. She would do her part to make this happen. By the time of her death in

[6] *Ann Clare Brokaw to Clare Boothe Luce, July 7, 1943, as quoted in Morris, 2014, pp. 45–46.*

[7] *Hatch chronicles the conversations between Father Sheen and Clare on pp. 176–185. Clare documented her own journey to Catholicism in "The Real Reason," an article that appeared in three installments in* McCall's *magazine in February, March and April, 1947.*

[8] *The close relationship Hank developed with Clare is everywhere evident in his tribute to her in "Clare Boothe Luce—Woman of the Century: A Son's Tribute,"* Crisis Magazine, *December 1, 1987.* https://www.crisismagazine.com/1987/clare-boothe-luce-woman-of-the-century-a-sons-tribute.

[9] *When responding to a query about an initial Visiting Assistant Professorship Program, Terrill Lautz, Program Officer of the Henry Luce Foundation, may have provided further insight into the overall aims of the still-to-be established Clare Boothe Luce Fund. "[T]he Luce Foundation wants to encourage the development of a permanent core of women faculty in fields where women have not been well represented in leadership positions at American universities [Lautz to Yu, 2 July, 1987]."*



qualified charitable organization, then to such other charitable or educational organization or organizations and in such shares as shall be determined by my Executors, on condition that such bequest be used exclusively for funding scholarships and professorships for women students and professors, in accordance with this paragraph FOUR.

Such amounts bequeathed to the HENRY LUCE FOUNDATION, INC., under this paragraph (or to such other organizations hereunder), shall be held and maintained by said Foundation as the CLARE BOOTHE LUCE FUND in accordance with the normal and usual procedures and policies of the Foundation, except that

dedicated exclusively to funding scholarships and professorships for women students and professors at educational institutions,

The purpose of the CLARE BOOTHE LUCE FUND shall be to encourage women to enter, study, graduate, and teach in the following fields of endeavor: Physics, Chemistry, Biology, Meteorology, Engineering (Electrical, Mechanical, Aeronautical, Civil, Nuclear and other Engineering disciplines), Computer Science, and Mathematics. I select such fields of endeavor in recognition that women today have already entered the fields of medicine, law, business and the arts, and in order to encourage more women to enter the fields of science.

- 12 -

a portion of the principal of such fund in the amount of THREE

THREE MILLION AND NO/100 DOLLARS ($3,000,000) each:

1. BOSTON UNIVERSITY (BOSTON, MASSACHUSETTS)
2. COLBY COLLEGE (WATERVILLE, MAINE)
3. CREIGHTON UNIVERSITY (OMAHA, NEBRASKA)
4. FORDHAM UNIVERSITY (NEW YORK, NEW YORK)
5. GEORGETOWN UNIVERSITY (WASHINGTON, D.C.)
6. MARYMOUNT COLLEGE (ARLINGTON, VIRGINIA)
7. MOUNT HOLYOKE COLLEGE (SOUTH HADLEY, MASSACHUSETTS)
8. MUNDELEIN COLLEGE (CHICAGO, ILLINOIS)
9. NOTRE DAME UNIVERSITY (SOUTH BEND, INDIANA)
10. SANTA CLARA UNIVERSITY (SANTA CLARA, CALIFORNIA)
11. ST. JOHN'S UNIVERSITY (LONG ISLAND, NEW YORK)
12. SETON HALL UNIVERSITY (SOUTH ORANGE NEW JERSEY)
13. TRINITY COLLEGE (WASHINGTON, D.C.)
14. VILLANOVA (OJAI, CALIFORNIA)

Physics, Chemistry, Biology, Meteorology, Engineering (Electrical, Mechanical, Aeronautical, Civil, Nuclear and other Engineering disciplines), Computer Science and Mathematics. All such scholarships and professorships shall be used exclusively for study in this country, not for travel or studies abroad.

- 13 -

Figure 2. Pages 12–13 of Clare Boothe Luce's last will and testament, noting her wish to fund women in STEM studies and careers.

1987, Clare had seen women make significant advances in some fields, but not, from her perspective, in mathematics, engineering, and certain sciences. Even though these disciplines fell outside her own areas of expertise, she recognized the need for support and funding. A comment late in her life may offer insight into Clare's choice for her legacy. In 1981, she admitted to one of her biographers, Wilfred Sheed, that she envied Sandra Day O'Connor, America's first female Supreme Court justice. Although Clare was a woman of many "firsts" herself, she told Sheed, "I don't want to be her, I [would] just like to have had that kind of chance [Sheed, p. 163]." Her creation of the Clare Boothe Luce Fund provided women with a *chance*.

She also publicly recognized the three most important breakthroughs for women that could help them achieve equal opportunity: the legal process, the female contraceptive, and opportunities for higher education [Luce, "Women Superior to Men," p. 281]. She especially valued higher education. Since Clare had only a very limited formal education, she was thrilled to witness—and advance—new opportunities for women in higher education. She was realistic, however. "Today she [a woman] is free to study for any 'masculine career' that her own ambition suggests" but… "as matters stand, her ambition is understandably dampened by the knowledge that even if she graduates at the top of her class, she will not find it easy to translate her well-earned degree into an upward-mobility job [Luce, 21st Century Woman, pp. 61–62]." Thus Clare understood that even though women had access to new educational opportunities, they still faced challenges in a male-dominated job market. Though Clare's success was not related to her level of education, she recognized that gender equality in education was a necessary precursor to job equality.

In her will, Clare designated "that the following named institutions shall be allocated a portion of the principal of such fund in the amount of…$3,000,000 each (about $6,692,000 today) [CBL Last Will and Testament, pp.12–13]." These schools included Boston University, Colby College, Creighton University, Fordham University, Georgetown University, Marymount College, Mount Holyoke College, Mundelein College, Notre Dame University, Santa Clara University, St. John's University (Long Island, NY), Seton Hall University, Trinity College (Washington, DC), and Villanova Preparatory School in Ojai California. [CBL Last Will and Testament, p. 13]. (Mundelein College affiliated with Loyola University Chicago in 1991 and is no longer funded from this initial allocation.) These fourteen schools had a Catholic affiliation, had awarded Clare an honorary degree, or had a sentimental attachment for her. Other schools can apply for funding through the "Clare Boothe Luce Program Invited Institution Competition." The Clare Boothe Luce Fund especially encourages Catholic institutions with strong science programs to apply [Clare Boothe Luce Program]. In 2017, by way of an example, eleven institutions received grants through the Invited Competition for funding to begin in 2018. Three decades after the initial bequest, in addition to the designated schools, more than 100 colleges and universities have benefited from the Clare Boothe Luce program [see CBL Recent Grants].

To fulfill Clare's aim "to encourage women to enter, study, graduate and teach," the Clare Boothe Luce Program administers awards in the three distinct categories of undergraduate scholarships and research awards; graduate fellowships for the first two years of a PhD program; and professorship support for the first five years of a tenure-track appointment. In the most recent year of funding for invited institutions, the Clare Boothe Luce Fund awarded grants in each of these categories [CBL Recent Grants].

Impact of a Designated Institution

Creighton University, one of the institutions designated in Clare's will, has a robust "Clare Boothe Luce Program for Women in Science [Creighton Clare Boothe Luce Program]." Through this program, Creighton funds undergraduate scholarships, graduate scholarships for women pursuing PhDs, and faculty positions. Since 1992, Creighton has rotated a Clare Boothe Luce Professorship in various fields in mathematics and science. Dr. Cynthia Farthing, who earned her PhD in mathematics from the University of Iowa, held the Clare Boothe Luce Professorship from 2007–2012. Dr. Catie Baker, an Assistant Professor in Computer Science, is currently the seventh Clare Boothe Luce Professor at Creighton. The Chair is designed to support a pre-tenure woman in a science or math field through tenure. It provides support to attend conferences, to fund undergraduate researchers, and to purchase supplies and materials.

Four of the six previous Clare Boothe Luce professors remain at Creighton and offer a strong network of support for Baker. Baker underscored the benefits of having a Clare Boothe Luce professorship at a designated institution. When she arrived at Creighton, she immediately shared a connection with the Clare Boothe Luce professors who had preceded her. This initial welcome segued into continued support and mentorship. At least two of the previous Clare Boothe Luce professors have moved into leadership positions on campus. These opportunities testify to the power of Clare's vision. Creighton's ongoing cycle of *chances* for women to earn a degree, teach others, and move into leadership positions is precisely the sort of outcome Clare aimed to achieve.

As part of her professorship, Baker oversees the selection of the Clare Boothe Luce undergraduate scholarships. Typically, Creighton offers 5–8 full tuition scholarships through the Clare Boothe Luce program. These scholarships are generally awarded to students who are actively engaged in research. Scholarship recipients take a "Women in Science" Seminar, taught by the Clare Boothe Luce Professor. This seminar focuses on issues facing women in science, including the impostor syndrome and stereotype threat.[10] The seminar fosters community and inspires conversations about graduate school, research, etc. As Dr. Baker described it, "the presence of the Clare Boothe Luce undergraduate scholarships creates an environment where women involved in undergraduate research are supported and valued [Interview with Catie Baker]." The ongoing presence of Clare Boothe Luce support at Creighton has not only advanced women at various stages in their careers but has also fostered a favorable environment on campus for women in mathematics and science to succeed as part of a broader community.[11]

Impact of Invited Institutions

Beyond the institutions designated in Clare's will, other eligible institutions of higher education can apply for awards through the "Invited Institution Competition." Sarah Spence Adams, Professor of Mathematics and Electrical & Computer Engineering at Olin College, received a Clare Boothe Luce scholarship for her final two years as a student at the University of Richmond in Richmond, Virginia in 1995–1997, for example. The award also included funding for undergraduate research. Her professor at Richmond, Dr. James Davis, called the Clare Boothe Luce opportunity to her attention and encouraged her to apply. At the time, she had no idea what "undergraduate

[10]*[Harris] includes a sample reading list for the Creighton Women in Science Seminar on pp. 109–110.*

[11]*For more on the impact of the early years of Clare Boothe Luce Funding at Creighton, see [Harris]. Although written in 1995, her insights apply to contemporary issues. As Harris puts it, "[a] topic of particular concern to students in the past 2 years has been sexual harassment…It is imperative that women not internalize harassment, whether it is called harassment or not. This is particularly true for gender-based harassment. Sexual harassment is much easier to identify, but gender-based harassment is far more common and more dangerous to the self-esteem and success of women. Examples of gender-based harassment include females being ignored in class (not called on) or, when they are called on, a female student's answer being deemed not as correct as a male student's identical response [Harris, p. 107]."*

research" meant. Davis showed her a book with an open question he had solved and helped her understand what undergraduate research might look like for her. She studied coding theory with Davis and cryptography with Dr. Gary Greenfield with her Clare Boothe Luce summer undergraduate research support.

The scholarship served as an "enormous source of confidence that I could actually be part of a mathematical research community," Adams says. "I wasn't exactly sure what that meant at the time but I understood that I had received funding to do mathematics. That was a novel idea." She presented her research at the Joint Mathematics Meetings in 1996 and won a prize for her poster. The prize was a gift certificate to select a book at a publisher. As she described it, "[to] claim the prize, I walked into the exhibit hall and was able to pick out any book I wanted. At that point, I only had books that my professors had assigned to me. Choosing my own mathematics book made me feel like a real mathematician [Interview with Sarah Spence Adams]."

Her undergraduate research experiences at Richmond made her a viable candidate for Joe Gallian's Research Experience for Undergraduates (REU) at the University of Minnesota in Duluth. She could take her Clare Boothe Luce funding with her so she did not have to rely on Gallian's NSF resources. The REU "propelled her into research" and helped her gain admission to the NSA Director's Summer Program the following year. These experiences not only improved her level of mathematics but also continued to open doors for her. She had the confidence to pursue a PhD in mathematics at Cornell, specializing in algebraic coding theory, and then to accept a faculty position at Olin College in Needham, Massachusetts, where she is now Professor of Mathematics and Electrical & Computer Engineering and a former Associate Dean of Faculty Affairs and Development. Adams says, "The Clare Boothe Luce experience taught me the value of undergraduate research so I dedicated myself to mentoring undergraduates at Olin." In her first decade at Olin, she mentored around 30 students, approximately 25 of whom continued for multiple years. All but three of these students have published professional journal articles with Adams. "I took their mentorship seriously," Adams explained. "I knew the impact it would have on them to come up with novel results, to edit, to revise, to publish, to attend a conference, to give a talk, to field questions, etc. I knew these values because I discovered them as an undergraduate myself [Interview with Sarah Spence Adams]."

Since Adams received her Clare Boothe Luce support more than twenty years ago, she provides an advantageous perspective on the long-term benefits of the program. "My Clare Boothe Luce experience was officially two years long. As the days, months and years have gone by, however, Adams notes, I have realized how much I gained from the scholarship and the opportunities that came along with it." Not surprisingly, when the Olin Development Office reached out to Adams to help craft an application for a Clare Boothe Luce grant to support undergraduate research, she was eager to help. Olin modeled their proposal around the two-year undergraduate research experience Adams had at Richmond. Olin's proposal included academic support, summer support, travel to conferences, and travel to see mentors. Adams recalls, "I had seen all of these components at Richmond and with Joe Gallian at Duluth and knew the impact they had on me." In 2011, Olin received an $180,000 award from the Clare Boothe Luce Foundation. Olin granted their first awards in 2013 [Bailey].

Epilogue

In February, 1942, Clare posed a question of possibility to her daughter Ann. "Would it amuse you," Clare asked, "to have your ma run for Congress and one day get to be a Cabinet minister, or maybe the first lady Vice President? [Morris, 1997, p. 473]." A year later, Albert P. Morano, Clare's executive assistant when she served as a Congresswoman, remarked that she "might even get to be President [Morris, 2014, p. 22]." Thus Clare and Morano at least considered the chance of Clare as the Vice President and/or President of the United States.[12] We know Clare valued a *chance*, for herself, and, as it turns out, for other women.

By the time Clare signed her will in early 1987, her experiences had more than acquainted her with the realities of life as an ambitious woman who exceeded the expectations of the existent social milieu. In perpetuity, then, she drew from these experiences to encourage women to pursue education for careers in fields, that at the time of her death, Clare viewed as primarily available to men. The last three decades testify to the continued vibrancy and veracity of her ideas.

Drawing from her two generations of experience with Clare Boothe Luce awards for mathematics, Sarah Spence Adams observed that "Clare Boothe Luce awards build confidence and create opportunities."[13] That formidable combination has advanced women not only in mathematics, but also in science and engineering, precisely what Clare Boothe Luce hoped to accomplish with her bequest and what the Clare Boothe Luce Fund aims to achieve today. Clare Boothe Luce may not have understood the intricacies of the fields she supported. She did, however, understand the necessary general framework for women to forge new

[12] *A decade after her death, Gore Vidal went so far as to say, "If born a man," she "could have easily been a president, for what that's worth these days: a cool billion, I believe." Vidal, p. 216.*

[13] *Of course, confidence is also a helpful skill for men in mathematics. University of Chicago mathematician Gilbert Ames Bliss noted the confidence his colleague, E. H. Moore, a pivotal figure in American mathematics in the late 1800s and early 1900s, acquired during his year of study in Berlin and Göttingen in 1885–1886. "There is no doubt," Bliss wrote, "that the year abroad affected greatly…Moore's career as a scholar. It established his* confidence *in his ability to take an honorable place in the … circle of mathematicians… [as quoted in Parshall and Rowe, p. 282, my emphasis]."*

pathways and find success. Clare Boothe Luce drew from her own experiences and observations to lay out the details for a foundation that would continue to promote and ensure these goals over time.

Bibliography

Unpublished Sources

Adams, Sarah Spence. Phone Interview. 8 August, 2018.

Baker, Catie. Phone Interview. 7 August, 2018.

Terrill E. Lautz to Yu Song-quing, July 2, 1987. Clare Boothe Luce Papers, Box 717, Folder 8, Clare Boothe Luce Papers. Manuscript Division. Library of Congress.

Clare Boothe Luce to Andrew Heiskell, February 4, 1968, p. 8. Clare Boothe Luce Papers, Box 716, Folder 3, Memorial Foundations, Clare Boothe Luce Papers. Manuscript Division. Library of Congress.

Clare Boothe Luce Last Will and Testament. Clare Boothe Luce Papers, Box 67, Folder 9, Clare Boothe Luce Papers. Manuscript Division. Library of Congress.

Published Sources

Bailey, Chelsea. More Opportunities for Women in Stem Research, September 27, 2013. www.olin.edu/blog/career-and-graduate-stories/post/more-opportunities-women-stem-research.

Brenner, Marie. *Great Dames: What I Learned from Older Women*. New York: Crown Publishers, 2000.

CBL Recent Grants to Invited Institutions. www.hluce.org/cblgrants.aspx.

Creighton Clare Boothe Luce Program for Women in Science. biology.creighton.edu/luce.

Grant Spotlight: The Clare Boothe Luce 25th Anniversary Professors Conference. www.hluce.org/cbl25thanniversaryspotlight.aspx.

Harris, Holly. "The Clare Boothe Luce Program at Creighton University," in *Teaching the Majority: Breaking the Gender Barrier in Science, Mathematics and Engineering*, pp. 98–110, edited by Sue Rosser, New York: Teachers College Press, 1995.

Hatch, Alden. *Ambassador Extraordinary: Clare Boothe Luce*. New York: Henry Holt and Company, 1955.

Krebs, Albin. Clare Boothe Luce Dies at 84: Playwright, Politician, Envoy. *New York Times*, October 10, 1987.

Luce, Clare Boothe. *Europe in the Spring*. New York: Alfred A. Knopf, Inc. 1940.

Luce, Clare Boothe. Talking Up—and Thinking Down: How to be a Success in Society without Saying a Single Word of Much Importance, *Vanity Fair* 1930: pp. 39, 85.

Luce, Clare Boothe. The Real Reason, parts 1–3. *McCall's*, February, March, April, 1947.

Luce, Clare Boothe. *The Women*. 1937. New York: Dramatists Play Service, Inc., 1966.

Luce, Clare Boothe. The 21st-Century Woman—Free at Last? *Saturday Review World* 24, August 1974: 58–62.

Luce, Clare Boothe. When Women Will Be Superior to Men, *McCall's*, April 1976: 186–187, 281–282.

Luce III, Henry. Clare Boothe Luce—Woman of the Century: A Son's Tribute, *Crisis Magazine*, December 1, 1987. https://www.crisismagazine.com/1987/clare-boothe-luce-woman-of-the-century-a-sons-tribute.

Martin, Ralph G. *Henry and Clare: An Intimate Portrait of the Luces*. New York: G. P. Putnam's Sons, 1991.

Morris, Sylvia Jukes. *Rage for Fame*. New York: Random House, Inc. 1997.

Morris, Sylvia Jukes. *Price of Fame: The Honorable Clare Boothe Luce*. New York: Random House, 2014.

Parshall, Karen Hunger & David E. Rowe. *The Emergence of the American Mathematical Research Community 1876–1900: J.J. Sylvester, Felix Klein, and E.H. Moore*. Providence, Rhode Island: The American Mathematical Society, 1994.

Sheed, Wilfrid. *Clare Boothe Luce*. New York: Dutton Publishing, Inc. 1982.

Teltsch, Kathleen. Mrs. Luce Left $70 Million for Women's Science Education, *The New York Times*, July 2, 1989. https://www.nytimes.com/1989/07/02/us/mrs-luce-left-70-million-for-women-s-science-education.html.

The Clare Boothe Luce Program. www.hluce.org/cblprogram.aspx.

US Department of Health Education and Welfare. "100 Years of Marriage and Divorce Rates in the United States, 1867–1967." Series 21, Number 24, December, 1973. https://www.cdc.gov/nchs/data/series/sr_21/sr21_024.pdf.

Vidal, Gore. *The Last Empire Essays 1992–2000*. New York: Doubleday, 2001.

Della Dumbaugh

Credits

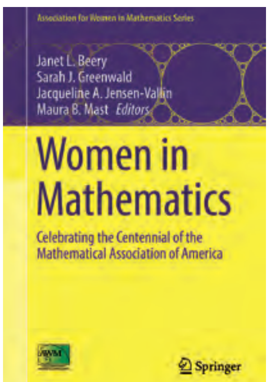
Figure 1 photo of Clare Boothe Luce by National Archives and Records Administration [Public domain], via Wikimedia Commons.

Figure 2 photos of select pages from Clare Booth Luce's last will and testament are courtesy of the Library of Congress.

Author photo used by permission of the University of Richmond.

Women in Mathematics

A Review by Katie Spurrier Quertermous



Women in Mathematics: Celebrating the Centennial of the Mathematical Association of America
Janet L. Beery, Sarah J. Greenwald, Jacqueline A. Jensen-Vallin, and Maura B. Mast, editors
Springer International Publishing, 2017
ISBN: 978-3319666938

In the last three years,[1] at least three mass market books—Margot Lee Shetterly's *Hidden Figures*, Nathalia Holt's *Rise of the Rocket Girls*, and Liza Mundy's *Code Girls*—and a major motion picture—*Hidden Figures*, based on Shetterly's book—have captivated audiences with the previously overlooked stories of women mathematicians who worked as human computers and cryptographers for the United States government. These works have offered the public a glimpse into the ongoing efforts of mathematicians and historians to write a social history of the lived experiences and contributions of women in the mathematical sciences.[2] *Women in Mathematics* adds to this literature with a collection of twenty-one engaging articles that include biographies, historical and cultural studies, and profiles of outreach and education initiatives related to women in mathematics. Most of the articles in this collection are written for a broad mathematical audience that includes students.

This volume grew out of a contributed paper session at MAA MathFest 2015 that was sponsored by the Association for Women in Mathematics (AWM). In connection with the celebration of the one hundredth anniversary of the Mathematical Association of America, the session sought to "recognize the contributions, achievements, and progress of women mathematicians over the past 100 years" through "talks about mathematics done by women and historical or biographical presentations celebrating women in mathematics." As the editors note in their preface, the resulting collection of articles is a mix of current scholarship and exposition on a wide variety of topics related to women in mathematics as opposed to a balanced study of the participation of women in mathematics during this time. Some of the articles summarize or extend work that has appeared previously, including Judy Green and Jeanne LaDuke's detailed documentary history of all of the American women who earned PhDs in mathematics from American and European universities between 1886 and 1939 and Margaret Murray's research on American women who earned PhDs in mathematics between the years 1940 and 1959. As a result, the volume also serves as a survey of a portion of the existing literature and compellingly invites the reader to delve deeper into that work.

The first two parts of the book are dedicated to telling the stories of women mathematicians in articles that range in style from formal historical and cultural studies to personal reflections and collections of interviews. These articles include more than eighty biographical profiles of women mathematicians and statisticians as well as numerous more concise descriptions of the experiences and contributions of women in these fields. The profiles are a mix of short sketches grouped within larger discussions of the mathematical and social context of a particular time, place, or culture and more in-depth studies of the professional and personal lives of individual women. Most of the profiles

---

Katie Spurrier Quertermous is an associate professor of mathematics and statistics at James Madison University. Her email address is querteks@jmu.edu.

[1]*This number of years is correct through April 5, 2019.*

[2]*Margaret Murray describes the concept of social history in more detail in Chapter 5 of* Women in Mathematics.

Communicated by Notices *Book Review Editor Stephan Ramon Garcia.*

DOI: http://dx.doi.org/10.1090/noti1861

focus on women from the United States, Canada, and Europe whose career paths are connected to academia or secondary education during at least a portion of their professional careers. Several chapters highlight women from underrepresented groups in mathematics, although the total number of biographical profiles of women from these groups is still small. The authors can be commended for including profiles of women whose stories are not widely known, so readers should expect to encounter at least a few unfamiliar names in these pages.

By presenting many of the profiles in groups, the articles emphasize both the connections between the individual women and the diversity of professional paths that they pursued even in times of limited career options. In their article on Girton College, Cambridge, Shawnee McMurran and James Tattersall profile a group of ten women who studied at Girton between 1880 and 1900 and achieved honors on the Mathematical Tripos exam despite the fact that women could not earn degrees from Cambridge at this time. These women applied their mathematical training to achieve success in numerous areas. As examples, Charlotte Angas Scott and Hertha Ayrton, the founders of Girton's Mathematical Club, had widely recognized research careers, Scott in algebraic geometry and Ayrton in engineering. Kate Knight Gale taught for many years, eventually becoming co-owner and joint headmistress of a school in South Africa, and Margaret Frances Evans was Mathematical Mistress of St Leonards School before ending her professional career to focus on family life. Emily Perrin and Beatrice Mabel Cave-Browne-Cave were both computers in Karl Pearson's statistical research lab at University College, London. In the 1930s, Girton College's Yarrow Research Fellowship supported the early work of Olga Taussky-Todd and Mary Lucy Cartwright, who both became prolific research mathematicians, Taussky-Todd in matrix theory and number theory and Cartwright in function theory and differential equations. Cartwright was elected Mistress of Girton College in 1948, and she led the college in this role for nineteen years while continuing her active involvement in the mathematical community.

Erica Walker explores the history of Black women in mathematics in the United States in a reflective essay that draws upon research she conducted for her 2014 book, *Beyond Banneker: Black Mathematicians and the Paths to Excellence*. As part of this essay, Walker juxtaposes the stories of Euphemia Lofton Haynes (PhD 1943) and Evelyn Boyd Granville (PhD 1949), the first two Black women to be awarded doctorates in mathematics in the United States.[3] Although they were born thirty-four years apart, both women were raised in Washington, DC, attended the same segregated high school, now called Dunbar High School, and earned undergraduate degrees from Smith College.


Euphemia Rosalie Lofton from Washington DC, Class Book, Smith College, 1914, Smith College Archives, Special Collections, Smith College.

Evelyn Boyd, Lawrence House, Smith College Yearbook, 1945, Smith College Archives, Special Collections, Smith College.

Figure 1. Euphemia Lofton Haynes (Smith College Class of 1914), left, and Evelyn Boyd Granville (Smith College Class of 1945), shown here in their college yearbook photos, attended the same Washington, DC, high school and the same college before earning PhDs in mathematics.

Dunbar was known for having committed, highly educated teachers who encouraged their students to attend college. The influence of these teachers seems to have played a strong role in both Haynes's and Granville's enrollments at Smith. Despite their many commonalities, Granville has said that she did not learn about Haynes until 1999, almost twenty years after Haynes's death. Haynes taught in the DC public school system, established the mathematics department at Miner Teachers College (now part of the University of the District of Columbia), served as president of the DC Board of Education, and was a key advocate for integration of the DC public schools. Granville's career in industry at IBM and other NASA contractors was bookended by academic positions, including faculty positions at Fisk University and the University of Texas at Tyler. Continuing the chain of influence begun by the teachers at Dunbar, two Black women taught by Granville, Etta Zuber Falconer and Vivienne Malone-Mayes, earned PhDs in mathematics and mentored additional generations of mathematicians.

The celebration of women's often under-recognized contributions in areas such as mentoring, teaching, professional service, academic administration, and advocacy is a theme that runs throughout most of the articles. It is impossible to categorize all of these contributions here, so I will just mention a few that I learned about for the first time while reading this book. An article by Emelie Agnes Kenney explores the vital roles that women played in keeping mathematics alive in Nazi-occupied Poland by studying and teaching in the clandestine education system despite the danger they faced if caught. Kenney recounts how Irena Gołąb disguised her math classes as crochet circles to avoid detection. Patti Hunter discusses Gertrude Cox's efforts to support statistics training around the world, efforts that are

[3]*Marjorie Lee Browne also finished the requirements for a PhD in mathematics in 1949, but her degree was not awarded until 1950.*

not as widely known as her achievements in building statistics programs in the United States or her service as president of the American Statistical Association. As part of this work, Cox was a program specialist at Cairo University's Institute of Statistical Studies and Research in 1964–1965. She was a strong advocate for the importance of statistical consulting and personally consulted on numerous projects in Cairo. Norma Hernandez earned a PhD in mathematics education in 1970 and was a faculty member at the University of Texas at El Paso for thirty years, serving as dean of the College of Education for six of those years. Hernandez was born and raised in El Paso, and Luis Ortiz-Franco investigates how Hernandez's life experiences in this multicultural city likely influenced her research on the relationships between culture and mathematics in the context of the K–12 mathematics education of Latinx students.

Woven throughout the historical accounts of women's contributions are discussions of some of the challenges the women faced in their pursuit of mathematical careers, especially those they encountered prior to the 1970s. Each woman's story is different, but common obstacles include barriers to advanced training, bias against women in hiring and promotion practices, and a lack of recognition of women's accomplishments. Many of these obstacles reflect the prevailing social norms of the women's times, and this context is important for helping readers understand the significance of individual and collective contributions as well as the dedication and perseverance of the women who made these contributions.

Multiple authors comment anecdotally on the shifts in mathematical culture that have occurred during their lifetimes, but a formal analysis of these changes is not the focus of this volume. Readers are, however, offered glimpses into a few of the advocacy efforts that championed changes in culture. Jacqueline Dewar describes outreach activities to encourage middle and high school girls to study mathematics that grew out of regional organizing meetings of the AWM during the 1970s. Laura Turner explores discussions at Canadian Mathematical Society meetings in the late 1980s and early 1990s that highlighted the underrepresentation of women on journal editorial boards and as plenary lecturers at meetings, and Sue Geller discusses skits presented at the Summer and Winter Joint Mathematics Meetings from 1990 to 1994 that used humor to draw attention to micro-inequities.

The third part of the book focuses on outreach and educational efforts. In a joint article, Jacqueline Dewar and Sarah Greenwald outline courses they have developed on women and mathematics that combine history, mathematical work, and equity issues and suggest opportunities for readers to experiment with these ideas through shorter-term outreach activities. Karl Schaffer describes the process of creating *Daughters of Hypatia*, a full-length dance performance that shares the stories of historical and contemporary women in mathematics, showcases mathematical thinking embedded in arts and crafts, and challenges stereotypes through dance and music. In an example of the connections between articles, *Daughters of Hypatia* includes an adaption of one of Sue Geller's skits on micro-inequities mentioned above. Sylvia Bozeman, Susan D'Agostino, and Rhonda Hughes discuss the EDGE Program (Enhancing Diversity in Graduate Education), which supports women mathematicians from diverse backgrounds through an annual summer session for beginning graduate students and ongoing mentoring networks. As part of this article, the authors describe aspects of the program that are designed to increase students' ability to successfully navigate the academic, cultural, and social transitions they will encounter in graduate school.

Figure 2. *Daughters of Hypatia* uses dance to share the stories of women mathematicians. In this 2015 photo, Laurel Shastri, left, and Lila Salhov perform "A Circle Has No Sides," one of several pieces that employ circular imagery.

Readers with a wide variety of mathematical, educational, and historical interests will find the articles in this collection engaging and inspiring. Most of the articles are accessible to undergraduate and graduate students although a faculty mentor may be required at times to help students contextualize the importance of contributions described in terms of journal titles and professional committees. The material is well suited for inclusion in existing courses and math club activities, and it provides opportunities not only to teach students about the history of women in mathematics but also to introduce them to important elements of mathematical culture through stories in which women play central roles. For example, Amy Shell-Gellasch's article on Mina Rees provides an excellent starting point for discussing both research funding and the value of conference attendance. Rees was instrumental in shaping federal funding of mathematics research through her work with the US government's Applied Mathematics Panel and the Office of Naval Research, and this article recounts, in Rees's own words, the importance she placed on the personal connections and broad understanding of