SURVEY LECTURE ON BILLIARDS

RICHARD EVAN SCHWARTZ

ABSTRACT

In this survey of billiards, I will discuss a variety of topics: rational polygonal billiards, irrational polygonal billiards, polygonal outer billiards, billiards in smooth ovals, and a bit about billiards in tables with scatterers.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 38C83; Secondary 32G15

KEYWORDS

Billiards, outer billiards, polygons, Riemann surfaces, geometry, dynamics

INTERNATIONAL CONGRESS OF MATHEMATICIANS 2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 4, pp. 2392–2429
DOI 10.4171/ICM2022/66

Published by EMS Press

1. INTRODUCTION

In one popular version of billiards, called *eightball*, the game is initialized with a triangular array of 10 polyester balls placed towards the back of a rectangular table that is about the length of a horse. The first player impacts the white *cue ball* with a tapered *pool stick*. The cue ball slides then rolls across green cloth, striking the other 10 balls and scattering them. This fateful start, which sets the game of eightball in motion, is called *the break*.

There are some who play billiards and there are some who think about billiards. Those who play care about the quality of the cloth, the weight of the balls, the feel of the stick. They grind the cushion that covers the front end of their pool stick into a tub of powdered chalk, trying to get just the right conditions of contact with the cue ball. Those who think about billiards usually free their minds from these physical properties and contemplate games of a more abstract nature.

This survey concerns the abstract games of mathematical billiards, henceforth simply called *billiards*. In billiards, the table might be a regular pentagon, or an ellipse, or the planar region bounded by a closed loop that is once but not twice differentiable. The tables might have obstacles in them, smaller shapes that the balls can bounce off as they move through the big table. Often there is just one ball in the game, a single point that slides without friction, but not always. There might be many balls, or charged particles influenced by magnetic fields. The game might be played on a sphere or in hyperbolic space.

Billiards is a huge, sprawling subject with deep connections to topics such as mathematical physics, ergodic theory, surface dynamics, Teichmüller theory, and algebraic geometry. There are many other surveys on billiards, most more comprehensive than this article. I will point some of these out later on. There would be no way for me to give a comprehensive survey of the whole business, even if I actually knew more than a little bit of it.

Rather, I will take the point of view that I am the proprietor of a Platonic pool hall. I built the establishment based on excellent advice but had to work quickly and on a limited budget. There are some beautiful rooms but also some of the plumbing and wiring is not quite up to code. Some doors lead nowhere at all. You have come to my establishment and I will show you around. I may entice you to stick around and play some games, maybe spend some money... The topics are organized mostly according to table type: square (Section 2), regular polygons (Section 3), rational polygons (Section 4), irrational polygons (Section 5), polygonal outer billiards (Section 6), convex ovals (Section 7), and tables with obstacles (Section 8).

This subdivision does not really cover it all, so sometimes I may drift off topic or omit important things. Some of the tables here, like rational polygons, are extremely crowded. You can barely hear yourself think above the shouting and the excitement. Other tables, like irrational triangles or convex ovals of intermediate differentiability, are much quieter. There are just a few patrons wandering around them and scratching their heads. I will also march you past the tables I have played on and (like it or not) regale you with tales of my own exploits from half-remembered “glory days.” My own work is concentrated in

Sections 5 and 6. If I had more space, I would also discuss magnetic billiards, Minkoswski billiards, symplectic billiards, polyhedral billiards, and billiards in hyperbolic space.

2. THE SQUARE

In the most common kind of billiards, a point moves along the table at unit speed and bounces off the sides according to the usual "angle in equals angle out" rule. The ball is not allowed to go into the corners. The square is the classic billiards table. I like the square $P = [0, 1/2]^2$ because it is nicely covered by the square torus $Y = \boldsymbol{R}^2/\boldsymbol{Z}^2$. There is a piecewise isometric map $f : Y \to P$, as indicated in Figure 1, which gives a bijection between geodesics (which miss the corners of the square tiling of Y) and billiard paths.

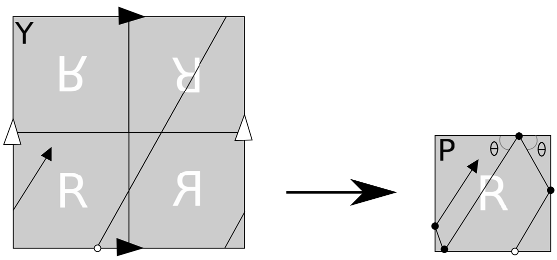

FIGURE 1
The torus covering the square.

2.1. Periodic billiard paths

A *periodic billiard path* is one that retraces itself. Each periodic billiard path on P corresponds to a closed geodesic on Y, which in turn corresponds to a line segment in the plane connecting $(0, 0)$ to some integer lattice point (m, n). There are infinitely many periodic billiard paths, but they come in a discrete set of maximal parallel families and there is one lattice point per family. The number $N'(L)$ of maximal parallel families consisting of periodic billiard paths of length at most L satisfies a beautiful asymptotic formula. Number $N'(L)$ counts the nonzero lattice points in the disk of radius L centered at $(0, 0)$, so

$$\lim_{L\to\infty} \frac{N'(L)}{L^2} = \pi. \tag{2.1}$$

How large is the error $E'(L) = N'(L) - \pi L^2$? A really crisp answer would be:

$$\lim_{L\to\infty} \frac{|E'(L)|}{L^{1/2}} = \infty, \quad \lim_{L\to\infty} \frac{E'(L)}{L^{(1/2)+\varepsilon}} = 0, \quad \forall \varepsilon > 0. \tag{2.2}$$

The first equation is a theorem proved independently by Hardy and Landau. The second equation is a famous open problem called *the Gauss Circle Problem*. See [56] for a survey.

The periodic billiard path is *primitive* if it does not trace several times over a smaller periodic billiard path. The lattice points (m, n) corresponding to primitive periodic billiard paths are coprime: m and n have no common divisors. Let $N(L)$ be the same count as above, but only for the primitive periodic billiard paths. We also ignore the orientation, which cuts the count in half. Estimating $N(L)$ recalls a happy exercise in number theory. The chances that a prime p does not divide (m,n) is $1-p^{-2}$. So, the proportion of coprime lattice points is asymptotically

$$\prod_p 1-p^{-2} = \left(\sum_{n=1}^{\infty} \frac{1}{n^2}\right)^{-1} = \frac{1}{\zeta(2)} = \frac{6}{\pi^2}.$$

This gives us

$$\lim_{L\to\infty} \frac{N(L)}{L^2} = \frac{\pi}{2\zeta(2)} = \frac{c_4}{\text{area}(P)}, \quad c_4 = \frac{3}{4\pi}. \tag{2.3}$$

2.2. Equidistribution

The aperiodic billiard paths on P correspond to geodesics γ on Y having irrational slope. Let γ_n denote the initial portion of γ having length n. We say that γ is *equidistributed* if, for all open $U \subset Y$,

$$\lim_{n\to\infty} \frac{\text{length}(\gamma_n \cap U)}{n} = \text{area}(U). \tag{2.4}$$

The following result establishes a dichotomy for square billiards. A geodesic is either closed or equidistributed.

Theorem 2.1. *Each irrational geodesic is equidistributed and hence dense.*

Proof. (Sketch.) The is equivalent to the statement that the orbits of an irrational rotation T of $\boldsymbol{R}/\boldsymbol{Z}$ are equidistributed in $\boldsymbol{R}/\boldsymbol{Z}$. That is, the fraction A_n/n converges to $|I|$, the length of I. Here A_n is the number of the first n orbit points contained in the interval I. Let I be an interval of length p/q. We can find powers $n_1,\dots,n_q$ such that the union

$$T^{n_1}(I) \cup \cdots \cup T^{n_q}(I)$$

covers $\boldsymbol{R}/\boldsymbol{Z}$ a total of p times, up to tiny overlaps and gaps that we can make as small as we like. Relatively speaking, very few orbit points fall into the tiny overlaps and gaps. So, by symmetry, $A_n/n \to p/q$. The case when $|I|$ is irrational follows from the case when $|I|$ is rational by a similar kind of limiting argument. ■

2.3. Connection to hyperbolic geometry

The group $\mathrm{SL}_2(\boldsymbol{Z})$ of integer 2×2 matrices of determinant 1 acts on $\boldsymbol{R}^2$ in such a way as to preserve $\boldsymbol{Z}^2$. Hence $\mathrm{SL}_2(\boldsymbol{Z})$ acts as affine automorphisms of Y. These maps permute the closed geodesics. One can study this action in terms of hyperbolic geometry. Let $\boldsymbol{H}^2$ denote the hyperbolic plane, given as the upper half-plane in $\boldsymbol{C}$. The *ideal boundary* of $\boldsymbol{H}^2$ is the extended real line $\boldsymbol{R}\cup\infty$. The group $\mathrm{SL}_2(\boldsymbol{Z})$ acts on $\boldsymbol{H}^2$ by the *linear fractional action*,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}(z) = \frac{az+b}{cz+d}. \tag{2.5}$$

When $z = \infty$, the right-hand side is set to a/c. The modular group, $\mathrm{SL}_2(\boldsymbol{Z})$, is an example of a *lattice*: $\boldsymbol{H}^2/\,\mathrm{SL}_2(\boldsymbol{Z})$ has finite hyperbolic area.

A *parabolic* element of the larger group $\mathrm{SL}_2(\boldsymbol{R})$ of real determinant-one matrices is one which is conjugate in $\mathrm{SL}_2(\boldsymbol{R})$ to an upper-triangular matrix. Such elements act on $\boldsymbol{H}^2$ without fixed points and they fix one point on the ideal boundary. The *cusps* of a subgroup of $\mathrm{SL}_2(\boldsymbol{R})$ are the fixed points of the parabolics. For $\mathrm{SL}_2(\boldsymbol{Z})$, the set of cusps is $\boldsymbol{Q} \cup \infty$. Thus, the periodic billiard directions are bijective with the cusps of $\mathrm{SL}_2(\boldsymbol{Z})$. The periodic direction of slope s corresponds to the cusp $1/s$.

To see more geometry of the correspondence, let us consider the closed geodesics on Y whose slopes lie in $[1, \infty]$. There is a familiar pattern of horodisks in $\boldsymbol{H}^2$ associated to the modular group. The horodisk associated to ∞ is the half-plane $\{z \mid \mathrm{Im}(z) \geq 1\}$. The horodisk tangent to the ideal boundary at $a/c \in [0, 1]$ is the round disk whose diameter is c^{-2}. This horodisk corresponds to the primitive closed geodesic whose slope is $c/a \in [1, \infty]$ and whose length is $\sqrt{a^2 + c^2} \in [c, 2c]$. Thus, the primitive closed geodesics of length about L and slope in $[1, \infty]$ correspond to those horodisks in $[0, 1]$ of diameter about L^{-2}.

2.4. Symbolic dynamics

Given an aperiodic billiard path in P, we can associate a biinfinite periodic binary sequence β. We record a 0 every time the path hits the horizontal side and a 1 every time the path hits the vertical side. Which sequences occur? I will explain the approach taken in [98] and also discussed in [27, 29, 100].

Call a biinfinite binary sequence *derivable* if either 00 never occurs or 11 never occurs in the sequence. Our sequence β is derivable. When the slope of the billiard path is less than (respectively greater than) 1, we never see 00 (respectively 11). If 00 does not occur, we let β' be the sequence obtained from β by removing a single 1 from every consecutive run of (1)s. The new *derived sequence* β' corresponds to the image of β under the element of $\mathrm{SL}_2(\boldsymbol{Z})$ which is the lower triangular matrix consisting of all 1s. In particular, β' is also derivable. We play the same game in the 11 case, with the roles of the digits swapped.

This analysis shows that the derivation process $\beta \to \beta' \to \beta'' \to \cdots$ can be continued forever, producing an infinite list of derivable sequences. Such sequences are called *Sturmian*. With just a few easily classifiable exceptions, every Sturmian sequences arises as the symbolic sequence for an irrational geodesic on Y. See [98].

3. REGULAR POLYGONS

Let P_n be the regular n-gon. To avoid exceptional cases, I will take $n \neq 3, 4, 6$. W. Veech [108, 109] noticed that many features of billiards on P_n resemble square billiards. At the end of the section, I will also explain a subtle dynamical difference.

3.1. The covering surface

As with the square, there is a surface X_n and an isometric map $f : X_n \to P_n$ which gives a bijection between geodesics on X_n and billiard paths on P_n.

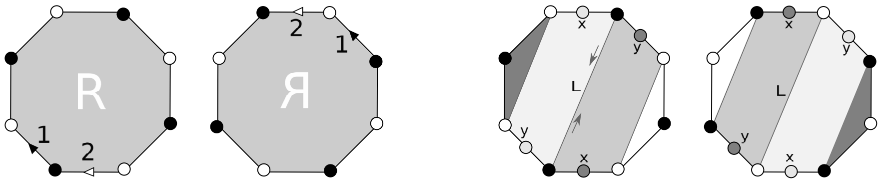

FIGURE 2
The octagon surface X_8.

The left-hand side of Figure 2 indicates the genus 3 surface X_8 made from the union of two octagons by gluing their sides together in the pattern indicated. The pattern is meant to continue to all 8 pairs of sides, and it is more natural if you think of the two octagons as stacked on top of each other in space. The white points are all identified and the black points are all identified. Away from these two special points of X_8, the surface is locally isometric to the plane. At each special *cone point*, X_8 is locally isometric to the space made by gluing together 8 sectors, each having angle $3\pi/4$. Each cone point has *cone angle* 6π. Our surface has a well-defined notion of direction, because the parallel rays pointing in any given direction in the plane induce a corresponding parallel vector field on X_8. The R-letters indicate the map $f : X_8 \to P_8$.

3.2. Connection to hyperbolic geometry

Let $A(X_n)$ denote the group of affine automorphisms of X_n. Such maps preserve the cone points and are locally affine away from them. There is a homomorphism

$$\phi : A(X_n) \to \mathrm{Isom}(\boldsymbol{H}^2). \tag{3.1}$$

Here $\phi(f)$ is defined to be the action of df, when df is interpreted as acting on $\boldsymbol{H}^2$. Because there is a well-defined notion of direction on X_n, there is a canonical identification of all the tangent spaces of X_n (away from the cone points). So, we can interpret df as acting on the same copy of $\boldsymbol{R}^2$ and then we get the hyperbolic action as above. Let

$$\Gamma(X_n) = \phi\big(A(X_n)\big). \tag{3.2}$$

This group is often called the *Veech group*, though sometimes one restricts to the orientation-preserving subgroup. Nontrivial affine maps of X_n which preserve the cusps cannot be too close to the identity, so $\Gamma(X_n)$ is a discrete group.

Theorem 3.1. *The group $\Gamma(X_n)$ is generated by reflections in the sides of a hyperbolic triangle with angles $0, 0, 2\pi/n$ when n is even and with angles $0, \pi/2, \pi/n$ when n is odd.*

Proof. (Sketch.) Consider the case $n = 8$. Simultaneous reflection in the vertical bisectors of our two octagons induces an affine automorphism corresponding to the hyperbolic reflection I_1 in the hyperbolic geodesic γ_1 connecting 0 to ∞. Likewise, simultaneous reflection in the bisector marked L on the right side of Figure 2 corresponds to an affine automorphism ι_2 and $I_2 = \phi(\iota_2)$ reflects in a geodesic γ_2 which makes an angle of $2\pi/8$ with γ_1.

The right-hand side of Figure 2 shows a decomposition of X_8 into 4 cylinders, all of the same modulus. On each of the big cylinders, there is an affine transformation, called a *Dehn twist*, which is the identity on the boundary and maps the vertex marked x to the same colored vertex marked y. The grey arrows show the motion of points near L. Thanks to the same-modulus condition, these maps extend to all of X_8, giving an affine automorphism of X_8 which restricts to Dehn twists on each cylinder. Call this map β. Let $\iota_3 = \iota_2 \circ \beta$ and $I_3 = \phi(\iota_3)$. Note that ι_3 has order 2 and preserves the directions parallel to the vertical and also parallel to L. So, I_3 is the reflection in the geodesic γ_3 connecting the appropriate endpoints of γ_1 and γ_2. The three geodesics bound the desired triangle.

The group Γ' generated by I_1, I_2, I_3 is the triangle group, and it is a subgroup of $\Gamma(X_n)$. The only discrete subgroup of $\mathrm{Isom}(\boldsymbol{H}^2)$ properly containing Γ' is the one generated by Γ' and the reflection in the bisector of our hyperbolic triangle. In the even case, this is not in $\Gamma(X_8)$, so $\Gamma' = \Gamma(X_8)$. In the odd case this extra element would be in the Veech group. ■

3.3. The Veech dichotomy

The Veech dichotomy [109] establishes the same kind of periodic/equistributed dichotomy for regular polygons that we saw for the square in Section 2.2.

Theorem 3.2. *(1) In each direction of X_n corresponding to a cusp of $\Gamma(X_n)$, there is a partition of X_n into metric cylinders foliated by parallel closed geodesics. (2) Conversely, any direction containing a closed geodesic corresponds to a cusp of $\Gamma(X_n)$. (3) Every geodesic in a noncusp direction is equidistributed.*

Proof. (First statement.) Let $X = X_n$. Let h be the parabolic affine automorphism of X corresponding to the cusp. Replacing h by h^2 if necessary we can assume that h fixes both cone points. Consider a geodesic ray γ emanating from one of the cone points in the direction fixed by h. Then h is the identity on γ. If γ does not return to a cone point then some small Euclidean disk $D \subset X$ intersect γ in at least 2 parallel strands. Inside D the map h acts as a shear and therefore shifts one strand of $\gamma \cap D$ with respect to the other in a nontrivial way. But h is the identity on both strands. This is a contradiction. Hence all geodesic rays emanating from the cone points return to cone points. The union of these *saddle connections* divides X into open cylinders foliated by parallel closed geodesics. ■

In the next section I will prove the second statement and a weak version of the third.

3.4. Periodic billiard paths

The Veech dichotomy relates the count of the periodic billiard paths to the enumeration of cusps in the Veech group. Veech [108, 109] makes this enumeration and proves that equation (2.3) holds for all regular polygons P_n with some constant c_n in place of c_4. The *Siegel–Veech constant* c_n in general is complicated, but here is the nice formula when respectively p is an odd prime and n is a power of 2:

$$c_p = \frac{p(p-1)(p^2+1)}{48(p-2)\pi}, \quad c_n = \frac{n^2(n-1)}{16(n-2)\pi}. \tag{3.3}$$

Veech's argument is a subtle mix of number theory and dynamics, but the horodisk picture discussed in Section 2.3 gives some geometric intuition. Rotate so that the horizontal direction is periodic. As with the modular group, consider a $\Gamma(X)$-invariant pattern of disjoint horodisks in $\boldsymbol{H}^2$, one per cusp. The primitive periodic directions of slope in $[1, \infty]$ and having length at most L essentially correspond to horodisks of diameter larger than about L^{-2} that are based at points in $[0, 1]$. There are certainly at most $O(L^2)$ of these horodisks, and it is at least plausible (and not too hard to prove) that there are at least $O(L^2)$ of them.

3.5. Symbolic dynamics and cusps

In [100], J. Smillie and C. Ulcigrai use the Veech group to show that the symbolic sequences for billiards on P_8 follow a derivation rule much like that for Sturmian sequences discussed above. Subsequently D. Davis [27] worked this out for all P_n.

In [73], A. Leutbecher proves the following result:

Theorem 3.3. *A point of the ideal boundary of the hyperbolic plane is a cusp of* $\Gamma(X_5)$ *if and only if it lies in* $\boldsymbol{Q}(\cos(2\pi/5)) \cup \infty$.

Similar results cover $n = 3, 4, 5, 6, 8$. Compare Theorem 1.5 in [80]. Is it an open problem to characterize the cusps of $\Gamma(X_n)$ for the cases other than these. The case $n = 7$ is the first unknown case. See [80] for a discussion of all this.

In [28], D. Davis and S. Lelièvre obtain many additional results about coding the billiards in P_5 and the cusps of $\Gamma(X_5)$.

3.6. Mixing

Here is one way billiards in P_n different from billiards in the square. On the square, the billiards map (which keeps track of the billiard paths at the bounce points) typically equidistributes the points, but it does not *mix them up*. A transformation T of a measure space (X, μ) is called respectively *mixing* and *weak mixing* if

$$\lim_{n\to\infty} \left|\mu\big(U \cap T^n(V)\big) - \mu(U)\mu(V)\right| = 0, \quad \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1}\left|\mu\big(U \cap T^j(V)\big) - \mu(U)\mu(V)\right| = 0. \tag{3.4}$$

A weak mixing transformation is "mixing at most times." On the square, the billiards map (see Section 7.1) is not weak mixing, but A. Avila and V. Delecroix [4] show that with respect to a typical aperiodic direction on P_n, the billiards map is weak mixing (but never mixing).

4. RATIONAL POLYGONS

A *rational polygon* is a polygon whose angles are all rational multiples of π. It is difficult to overstate (and to adequately survey) the spectacular development of rational billiards, the study of billiards in rational polygons and more generally the straight line flow in translation surfaces. For additional sources, see [46,77,115,118,119].

4.1. Translation surfaces

A *translation surface* is any oriented surface made by gluing together a finite collection of disjoint polygons in such a way that the side identifications are given by translations. As above, such a surface is locally Euclidean and has a well-defined sense of direction away from a finite number of cone points whose angle is an integer multiple of 2π. The geodesics on a translation surface are locally straight lines and they avoid the cone points. Here are 3 basic definitions.

Strata. The set of all translation surfaces with a fixed topological type and a fixed list of cone angles is called a *stratum*.

Veech group. The affine automorphism group $A(X)$ and the *Veech group* $\Gamma(X) = \phi(A(X))$ are defined for a general translation surface X just as in Section 3.2. We let $A_+(X)$ and $\Gamma_+(X)$ denote the respective orientation preserving subgroups. We sometimes think of $\Gamma_+(X)$ as a subgroup of $\mathrm{SL}_2(\boldsymbol{R})$ rather than as a subgroup of $\mathrm{Isom}(\boldsymbol{H}^2)$.

Lattice property. Group $\Gamma(X)$ need not be a lattice in $\mathrm{Isom}(\boldsymbol{H}^2)$, but when it is a lattice we say that X *has the lattice property*.

Lemma 4.1 (Katok–Zemylakov construction). *Let P be rational polygon. Then there is a translation surface X_P and a piecewise isometric map $f : X_P \to P$ which carries geodesics on X_P to billiard paths on P.*

Proof. For each edge e of P there is a reflection R_e in the line through the origin parallel to e. The group G generated by these reflections is finite, thanks to the rationality assumption. For each $g \in G$, define $P_g = g(P) + V_g$. Here V_g is a vector included so that all the polygons $\{P_g \mid g \in G\}$ are disjoint. The final answer is independent of these auxiliary vectors.

We form an identification space on the union of these polygons by gluing together every pair of edges of the form

$$e_1 = g(e) + V_g, \quad e_2 = gr(e) + V_{gr}, \quad r = R_e \tag{4.1}$$

by a translation. By construction, X_P is a translation surface. The map $X_P \to P$ is defined to be the inverse of the map $g + V_g$ on the piece P_g. ∎

This all-purpose construction is not necessarily the most efficient one in special situations. For instance, when applied to the regular n-gon it produces a cyclic cover of the surface X_n considered in the previous chapter.

4.2. Affine action

Each stratum has an action of $\mathrm{GL}_2(\boldsymbol{R})$ on it. If we have a translation surface Y made by gluing together a finite list $P_1, \ldots, P_k$ of polygons and an element $g \in \mathrm{GL}_2(\boldsymbol{R})$, we get a new translation surface $g(Y)$ by gluing together $g(P_1), \ldots, g(P_k)$ in the same pattern.

When Y is the square torus, and $g \in SL_2(\boldsymbol{Z})$, the surface $g(Y)$ is obtained by gluing the opposite sides of an area 1 parallelogram whose vertices have integer coordinates. This is just Y again, presented differently. More generally, if $g \in A_+(Y)$ then $dg(Y)$ is the same surface as Y. Conversely, $g \in SL_2(\boldsymbol{R})$ and $g(Y) = Y$ then $g \in \Gamma_+(Y)$. This lets us identity the orbit $SL_2(\boldsymbol{R}) \cdot Y$ with $SL_2(\boldsymbol{R})/\Gamma_+(Y)$. We may identify this latter space with the (orbifold) unit tangent bundle of $\boldsymbol{H}^2/\Gamma_+(X)$. Equipped with this point of view, let us prove the second statement of the Veech dichotomy theorem.

Lemma 4.2. *Suppose that X is a translation surface and $\Gamma_+(X)$ is a lattice in $SL_2(\boldsymbol{R})$. If X has a closed geodesic with direction δ, then δ corresponds to a cusp of $\Gamma(X)$.*

Proof. For ease of exposition, assume that elements of $\Gamma_+(X)$ act on $\boldsymbol{H}^2$ without fixed points, so that the quotient $\Sigma = \boldsymbol{H}^2/\Gamma_+(X)$ is a finite-area hyperbolic surface. In the general case, we would pass to a finite-index subgroup. A geodesic ray in Σ either goes to a cusp or else recurs infinitely often to a compact subset. We rotate X so that δ is horizontal. Let

$$g_t = \begin{bmatrix} e^{-t} & 0 \\ 0 & e^t \end{bmatrix}. \tag{4.2}$$

Since $g_t(X)$ has a closed loop whose length tends to 0 as $t \to \infty$, the surface $g_t(X)$ exits every compact subset of $SL_2(\boldsymbol{R})/\Gamma_+(X)$ as $t \to \infty$. But the set $\{g_t(X) \mid t \geq 0\}$ projects to a geodesic ray on Σ. If ∞ (the point on the ideal boundary corresponding to the horizontal direction) is not a cusp of Σ then this ray recurs infinitely often to a compact subset of Σ. But this is a contradiction. ∎

Here is a weaker version of the third statement. The reason that the aperiodic directions are dense is that the boundary of the complement of a nondense geodesic would have a closed loop, and then Lemma 4.2 would say that the direction corresponds to a cusp. The equidistribution statement follows from Theorem 1.1 in H. Masur's paper [76] and the fact that a geodesic which does not exit the cusp of $\boldsymbol{H}^2/\Gamma(C)$ is recurrent.

4.3. Connection to Teichmüller space

The space $\boldsymbol{M}_g$ is the space of Riemann surfaces of genus g. The universal orbifold cover of $\boldsymbol{M}_g$ is called *Teichmüller space* and denoted $\boldsymbol{T}_g$. To define $\boldsymbol{T}_g$, we fix a background genus g surface Σ_0. A point in $\boldsymbol{T}_g$ is then an equivalence class of pairs (Σ, ψ), where Σ is a genus g Riemann surface and $\psi : \Sigma_0 \to \Sigma$ is a homeomorphism. Two pairs (Σ_1, ψ_1) and (Σ_2, ψ_2) are equivalent if there is a biholomorphic map $f : \Sigma_1 \to \Sigma_2$ such that $f \circ \psi_1$ and ψ_2 are homotopic maps. The map ψ is often called the *marking* of Σ.

One can think of a genus g translation surface as a Riemann surface equipped with a holomorphic 1-form, an expression which looks like $f(z)dz$ in local coordinates. These objects are also called *abelian differentials*. Thus a translation surface Y naturally gives rise to a point in the "vector bundle" of abelian differentials over moduli space. I put "vector bundle" in quotes just because $\boldsymbol{M}_g$ is an orbifold rather than a manifold. More simply, each marked translation surface corresponds to a point in the vector bundle of abelian differentials over Teichmüller space.

We can interpret the $\mathrm{SL}_2(\boldsymbol{R})$ action as giving a group action on the abelian differential bundle over Teichmüller space. For any translation surface Y, the orbit $\mathrm{SL}_2(\boldsymbol{R}) \cdot Y$ projects to the hyperbolic plane in $\boldsymbol{T}_g$. When Y has the lattice property, this hyperbolic plane further projects to an isometric copy of $\boldsymbol{H}^2/\Gamma_+(Y)$ in $\boldsymbol{M}_g$ called a *Teichmüller curve*.

4.4. Structure of strata

The following lemma gives the basic structure of the strata.

Lemma 4.3. *A stratum having genus g and v cone points is a complex orbifold of (complex) dimension $n = 2g + v - 1$. The manifold cover of the orbifold has an atlas of coordinate charts with transition functions in* $\mathrm{GL}_n(\boldsymbol{C})$.

Proof. (Sketch.) Let Σ be the stratum and let $X \in \Sigma$ be a translation surface. We construct local coordinates called *period coordinates* to describe a neighborhood of X in Σ. Triangulate X so that the v vertices of the triangulation are the cone points. This realizes X as the quotient of a union of f triangles with e pairs of edges glued together by translations. Orient each pair of edges and pick one edge from each pair and record the complex number that describes its direction and magnitude.

A nearby assignment of e complex numbers gives a recipe for a new collection of triangles provided that, around each triangle, the corresponding sum of the complex numbers (perhaps with signs in front) is 0. This gives f relations, but one relation is redundant because the closing conditions on all but one triangle determine the last closing condition. So, the space of valid choices has dimension $n = d - f + 1 = 2g - v + 1$.

These coordinates might not give a local homeomorphism into $\boldsymbol{C}^n$ because (thanks to symmetries) different assignments can give rise to the same translation surface. By considering the same marking trick as with the definition of Teichmüller space, you can construct a cover of the stratum which is a manifold and for which the above coordinates are a coordinate chart in the usual sense. If X is triangulated in a different way then the transition functions between the coordinate charts are complex linear. ∎

Lemma 4.3 has a kinship with W. Thurston's paper [105], in which he considers the moduli space of flat cone metrics on the sphere with $n + 2 \geq 4$ prescribed cone angles. He constructs "period coordinates" in which these spaces are orbifolds whose transitions functions lie in $\mathrm{PU}(1, n-1) \subset \mathrm{GL}_n(\boldsymbol{C})$. Under certain arithmetic conditions on the cone points, Thurston shows that the subset of unit area structures is open dense in a complex hyperbolic orbifold coming from a Deligne–Mostow [31] lattice.

4.5. Periodic billiard paths

M. Boshernitzan, G. Galperin, T. Krüger, and S. Troubetzkoy [14] prove that the periodic billiard positions and directions are dense in a rational polygon.

Theorem 4.4. *For any rational polygon P, the set of periodic billiard paths lifts to a dense subset of the unit tangent bundle of P.*

The closed geodesics on a translation surface Y come in parallel families and sweep out maximal metric cylinders. Let $N(L,Y)$ denote the number of these maximal cylinders of length less than L. Here is a result of H. Masur's [74,75]:

$$0 < A_Y \le \liminf_{L\to\infty} \frac{N(L,Y)}{L^2} \le \limsup_{L\to\infty} \frac{N(L,Y)}{L^2} \le B_Y < \infty. \tag{4.3}$$

This result implies a similar-looking result for periodic billiard paths on rational polygons.

Here is one of the main open problems in the field.

Conjecture 4.5. *$A_Y = B_Y$ for all translation surfaces Y.*

A. Eskin and Masur [42] prove Conjecture 4.5 for almost all surfaces within each stratum, and they show that the common value, called the *Siegel–Veech constant*, just depends on the stratum. One application of the main result in [43, 44] (discussed below) is that Conjecture 4.5 is true in an average sense,

$$\lim_{L\to\infty} \frac{1}{L} \int_0^L N(Y, e^t) e^{-2t}\, dt = C_Y, \tag{4.4}$$

for all surfaces Y. The constant depends on the surface.

These strong asymptotic counting results rely on powerful dynamical results about the action of $SL_2(\boldsymbol{R})$ and its subgroups on the bundle of abelian differentials on Teichmuller space. This survey does not really touch on these ideas. See [115] for details.

4.6. Classification problem

Which polygons and translation surfaces have the lattice property? Which Teichmüller curves arise? How does the Teichmüller curve depend on the translation surface (with the lattice property)? Here is some progress on these questions.

Genus 2. In the genus 2 stratum with 2 cone points, only the regular decagon with opposite sides identified has the lattice property. See [79]. In the genus 2 stratum with 1 cone point, there is an infinite family, all coming from L-shaped polygons in which a rectangle is cut out the corner of the square. The corresponding Veech groups are classified by an invariant (± 1), called the *spin*, and square-free integer D congruent to 0 or 1 mod 4 called the *discriminant*. The corresponding quotient $\Sigma_D = \boldsymbol{H}^2/\Gamma(X_D)$ embeds in a Hilbert modular surface $(\boldsymbol{H}^2 \times \boldsymbol{H}^2)/\, PSL_2(O_D)$. Here O_D is the ring of algebraic integers in $\boldsymbol{Q}(\sqrt{D})$. See [78] and [20]. The topological features of Σ_D, namely its Euler characteristic (M. Bainbridge [6]) and orbifold points (R. Mukamel [85]), are known.

Genus 3 and 4. C. McMullen, R. Mukamel, and A. Wright [81] recently discovered new infinite families of primitive translation surfaces, i.e., not covers of other translation surfaces in lower genus, in genus 3 and 4 which have the lattice property. These surfaces correspond to certain dart-shaped quadrilaterals. One family corresponds to darts of the form $(1,1,1,9)$, and the other corresponds to darts of the form $(1,1,2,8)$. These numbers describe the relative proportions of the interior angles of the darts.

Triangle groups. The Veech groups all have cusps. With Theorem 3.1 in mind, one can ask whether every triangle group with a cusp arises as a Veech group. Subject to certain congruence conditions and an index 2 ambiguity, this turns out to be true. I. Bouw and M. Möller [15] discovered translation surfaces with this property and defined them in terms of algebraic geometry. Later, W. Hooper gave concrete combinatorial descriptions for them [61]. Compare also [29] and [116].

Computation. J. Athreya, D. Aulicino, and W. Hooper [3] explicitly compute the quotient $\boldsymbol{H}^2/\Gamma(X)$ where X is the translation surface associated to the regular dodecahedron. This canonical object has genus 131, 19 cone singularities, and 362 cusps. Relatedly, the authors show that modulo the action of $A(X)$ there are exactly 31 closed geodesics on X which contain just a single cone point. The point of these results is not just the surprising complexity, but also their vital reliance on software. In this case, the results use Sage-based software written by Hooper and V. Delacroix [30]. See [16] and [41] for computational Veech group algorithms.

4.7. Orbit closures

Let g_t be as in equation (4.2). When Y has the lattice property, the flow $t \to g_t(Y)$ traces out the lift of a geodesic to the unit tangent bundle U of $\boldsymbol{H}^2/\Gamma(Y)$. The closure of this set could be the lift to U_1 of a geodesic lamination, a set which looks locally like the product of a geodesic and a Cantor set. When Y does not have the lattice property, the orbit closure could be even wilder. A huge breakthrough in rational billiards rules out this kind of wildness for the closure of $\mathrm{GL}_2(\boldsymbol{R}) \cdot Y$. The following result is a combination of results in A. Eskin and M. Mirzakhani [43] and Eskin–Mirzakhani–A. Mohammadi [44].

Theorem 4.6. *For any translation surface Y, the closure of $\mathrm{GL}_2(\boldsymbol{R}) \cdot Y$ is a locally affine orbifold (possibly with self-intersections). The support of any ergodic $\mathrm{SL}_2(\boldsymbol{R})$-invariant measure on the orbit closure has the same structure, and the measure is affine.*

Theorem 4.6, which A. Zorich [119] calls the *Magic Wand Theorem*, has spurred a huge amount of activity. I have already mentioned equation (4.4) as an application. Here are three more developments.

Algebraic structure. The bundle of abelian differentials over moduli space has an algebraic structure, but the Magic Wand Theorem only says that the orbit closures are real analytic in this structure. (The periodic coordinates are transcendental.) In [45], S. Filip shows that nonetheless the orbit closures are algebraic.

The illumination problem. Given a billiard table P and two points $x, y \in P$, one says that *x illuminates y* if there is a billiard path starting at x and containing y. In [72], S. Lelièvre, T. Monteil, and B. Weiss use the Magic Wand Theorem to prove that for any rational polygon P and for any point $x \in P$, there are at most finitely many $y \in P$ such that x does not illuminate y. Compare also [1]. In a recent refinement [114], A. Wolecki shows that there are at most finitely many *pairs* in any rational polygon P such that x does not illuminate y.

Wild horocyclic closures. Let P be the parabolic subgroup consisting of the upper-triangular matrices. J. Chaika, J. Smillie, and B. Weiss [22] prove that the closure of $P \cdot Y$ can be wild. For instance, it can have fractional Hausdorff dimension. The surprise here is that when Y has the lattice property, the closure $P \cdot Y$ is either a single curve or all of U.

5. IRRATIONAL POLYGONS

Now we leave the excitement of rational billiards and consider the case of billiards on irrational polygons, those whose angles are not all rational multiples of π. Given the incredible depth and precision of the results about rational billiards, it is remarkable that we do not even know if every obtuse triangle has a periodic billiard path! We also do not know if every acute triangle has a periodic billiard path that is not of the kind shown in Figure 3 below. Are there polygons without periodic billiard paths? Nobody knows. Problems such as the illumination problem discussed above are wide open.

5.1. Easy examples

Figure 3 shows periodic billiard paths which exist on all acute triangles and all right triangles, respectively. The periodic path with 6 bounces shown on the left-hand side of Figure 3 is part of an infinite parallel family of such paths. This family degenerates to the periodic billiard path having 3 bounces. This special periodic billiard path, called the *Fagnano path*, is the inscribed triangle of minimum length. The billiard path on the right starts and ends perpendicular to the side. Call such periodic billiard paths *orthogonal*.

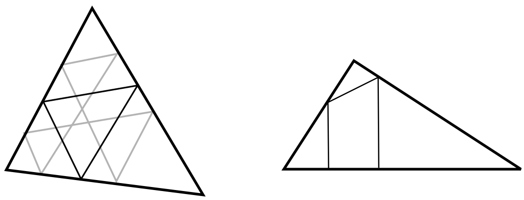

FIGURE 3
Periodic billiard paths on acute and right triangles.

5.2. Right triangles

When applied to irrational polygons, Lemma 4.1 produces an infinite translation surface. When P is a right triangle, the resulting surface X_P is really neat. Let Q denote the rhombus that is tiled by 4 copies of P. Now glue a countable collection of copies of Q around a single vertex, in a kind of spiral pattern, as indicated in Figure 4.

Finally, glue together the remaining sides of the infinite union in the pattern indicated. This surface is constructed in the abstract so that the different rhombi do not really lie in the plane and overlap. Surface X_P has 4 infinite cone points, all of which have the same

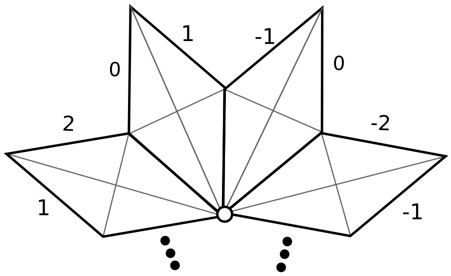

FIGURE 4
The translation surface associated to an irrational right triangle.

structure. The construction above favors one of the cone points, and so you have to stare at the picture for a while to see the other three.

In [107], S. Troubetzkoy analyzes these surfaces and proves the following result:

Theorem 5.1. *Any irrational right triangle has a side such that all but countably many points in that side are the start points of orthogonal periodic billiard paths.*

Compare the work of B. Cipra, R. Hanson, and A. Kolan [26]. A quick corollary of Theorem 5.1 is that periodic billiard paths on a right triangle are dense, as in Theorem 4.4.

A periodic billiard path on a triangle is called *stable* if a periodic billiard path of the same combinatorial type exists on all nearby triangles. W. Hooper [58, 60] proved the following result:

Theorem 5.2. *No combinatorial type of periodic billiard path exists on both acute and obtuse triangles. In particular, periodic billiard paths on right triangles are unstable.*

There are still lots of open problems about right triangular billiards. For instance,

Question 5.3. Does the number of maximal families of periodic billiard paths in a right triangle have quadratic growth, as in Masur's theorem?

5.3. Existence results for obtuse triangles

In [49], G. A. Galperin, A. M. Stepin, and Y. B. Vorobets construct some infinite families of periodic billiard paths in irrational polygons. In [57], L. Halbeisen and N. Hungerbuhler construct additional infinite families of periodic billiard paths in obtuse triangles. The examples in [49] and [57] are stable.

Here is one of my results, proved in a series of two papers [92, 94].

Theorem 5.4. *An obtuse triangle has a stable periodic billiard path provided all its angles are at most* 100 *degrees.*

Proof. (Rough sketch.) Let $\boldsymbol{P}$ denote the parameter space of similarity classes of obtuse triangles. Then $\boldsymbol{P}$ itself is a triangle. Let $\boldsymbol{P}_{100}$ denote the subset corresponding to triangles having less than 100 degrees.

Suppose that w is some finite word that corresponds to a stable periodic billiard path on some triangle. Let $O(w) \subset \boldsymbol{P}$ be the set corresponding to triangles which support a periodic billiard path with sequence w. We call $O(w)$ the *orbit tile*. To estimate $O(w)$, we successively reflect the initial triangle according to the digits of the word, as in Figure 5. The result is called the *unfolding*. The stability condition guarantees that the first and last sides of the unfolding are parallel, no matter which triangle we use for the construction. Rotate so that the translation carrying the first side to the last side is horizontal.

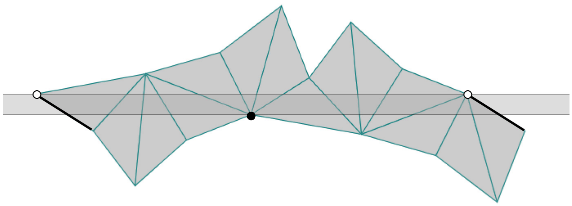

FIGURE 5
An unfolding and a corridor.

There is a maximal strip, with horizontal sides, such that any horizontal line in the strip corresponds to a periodic billiard path. We call this a *corridor*. As the triangle changes shape, the corridor widens or narrows according to how the vertices move. The billiard path disappears when the height of one of the vertices along the top of the unfolding has the same height as one of the vertices along the bottom. This condition is given by the vanishing of a finite trigonometric sum. Using some mixture of analytic and numerical methods, one can approximate $O(w)$ in a rigorous way.

Showing that a given region in $\boldsymbol{P}$ consists entirely of triangles with stable periodic billiard paths amounts to proving that one can cover this region with orbit tiles. First, I found some nice infinite families of orbit tiles which in a systematic way cover certain regions of $\boldsymbol{P}_{100}$, *trouble spots*, which have no finite cover. (See the discussion and also Theorem 5.8 below.) Then I do repeated depth-first searches through the tree of words, up to a suitable depth, in order to cover the remainder $\boldsymbol{P}_{100}$ with about 200 more orbit tiles. ∎

W. Hooper and I wrote a computer program, called *McBilliards*, which does these searches and also plots the orbit tiles to a high degree of accuracy. The search looks like it is exponential in the depth—which would be very bad—but in fact it is much faster. Given a triangle and a word, McBilliards performs the unfolding until it appears that the unfolding is so crooked that no continuation of the word would produce a nonempty corridor. This pruning vastly increases the speed.

Recently, in a preprint [106], J. Garber, B. Marinov, K. Moore, and G. Tokarsky improved my 100-degree result to 112.3 degrees, though they do not have the stability conclusion. They discovered that certain special triangles are quite difficult to cover with stable orbit tiles, but nonetheless have unstable periodic billiard paths. As I mentioned above, my dataset involved several infinite families of words and about 200 additional "sporadic" words. The dataset in [106] involved bazillions of sporadic words. I do not know the number, but it took me several hours just to download the dataset from the internet!

One serious difficulty in using the orbit tile approach to prove that every triangle has a periodic billiard path is that regions near the boundary of the parameter space, corresponding to thin triangles, are extremely hard to cover with orbit tiles. In my 100-degree theorem, I got lucky and found an infinite family of periodic billiard paths which cover a neighborhood of the boundary, up to $5\pi/8$. Beyond this, there is no known point on the boundary which has a neighborhood covered by orbit tiles. Note that $5\pi/8$ radians is 112.5 degrees, so the 112.3 result of [106] cannot be much improved without more luck at the boundary.

Here are a few more existence results. In [62], Hooper and I used similar ideas to prove the following result.

Theorem 5.5. *If $\{P_n\}$ is any sequence of triangles converging to an isosceles triangle, then P_n has a periodic billiard path once n is sufficiently large.*

There is only one counting result, due to Hooper [59], which even vaguely resembles equation (2.3).

Theorem 5.6. *There exists an open subset of obtuse triangles such that for each triangle in the set the number $N(L)$ of primitive periodic billiard paths has the property that*

$$\lim_{L\to\infty} \frac{N(L)}{L \log(L)^k} = \infty$$

for any k.

Numerical experiments with McBilliards lead to the following conjecture:

Conjecture 5.7. *Orbit tiles are connected and simply connected.*

It would also be interesting to know how the area of the orbit tile depends on the length of the word. One approach to showing that some triangles do not have any periodic billiard paths would be to show that in general the area decays very rapidly and the number of words does not grow quickly. See the result of D. Scheglov discussed below in Section 5.6.

5.4. Recalcitrance

Call a triangle T *recalcitrant* if for any $\varepsilon > 0$ there are triangles within ε of T (in terms of angle differences) supporting no periodic billiard paths of length less than $1/\varepsilon$. The corresponding point in the parameter space has no neighborhood covered by finitely many orbit tiles. In [92] I proved the following result:

Theorem 5.8. *The $(2, 3, 6)$ right triangle is recalcitrant.*

Theorems 5.8 and 5.4 complement each other. Basically, Theorem 5.8 says that a result like Theorem 5.4 is intrinsically hard. A neighborhood of the $(2,3,6)$ triangle, on the obtuse side of the parameter space, is one of the trouble spots I mentioned above in connection with the proof of Theorem 5.4.

Numerical experiments with McBilliards lead to the following conjectures:

Conjecture 5.9. *Every obtuse Veech triangle is recalcitrant.*

Conjecture 5.10. *Once the Fagnano orbit tile is removed, the acute triangle parameter space is not covered by any finite union of orbit tiles.*

Conjecture 5.11. *For any N there is some $\varepsilon > 0$ so that no triangle within ε of the equilateral triangle has an orthogonal periodic billiard path of length less than N.*

5.5. Bounce rigidity

One of the few sweeping geometric results about billiards in any polygon is *bounce rigidity*. Every polygon P gives rise to a collection $B(P)$ of biinfinite words corresponding to biinfinite billiard paths. These billiard paths may or may not be periodic. The set $B(P)$ is called the *bounce spectrum*. In [40], M. Duchin, V. Erlandsson, C. Leininger, and C. Sadanand prove that the bounce spectrum essentially determines the shape of P.

Theorem 5.12. *If two polygons P_1, P_2 are such that $B(P_1) = B(P_2)$, then either P_1 and P_2 are related by a similarity or else P_1 and P_2 have all right angles and are affinely equivalent.*

A very similar result is proved independently by A. Calderon, S. Coles, D. Davis, J. Lanier, and A. Oliveira in [19]. These results are the culmination of many works on this topic. See [40] and [19] for further references.

5.6. Ergodicity and complexity

A. Katok [68] has called the ergodicity and orbit growth for irrational polygonal billiards one of the five most resistent problems in dynamics. Here are two subproblems of Problem 3 on Katok's list.

Question 5.13. Is the billiard flow ergodic with respect to almost every polygon? What about with respect to almost every triangle?

Conjecture 5.14. *With respect to any polygon the number $S(L)$ of saddle connections of length less than L is eventually less than $L^{2+\varepsilon}$ for any $\varepsilon > 0$.*

The work of S. Kerkhoff, H. Masur, and J. Smillie [69] gives a G_δ-set of ergodic tables. Recently, J. Chaika and G. Forni [21] proved a similar result about weak mixing. Compare [5]. Ya. Vorobets [112] gives an explicit (but crazily impractical) criterion for ergodicity:

Theorem 5.15. *If the polygon Q admits approximation by rational polygons at the rate*

$$\phi(N) = \left(2^{2^{2^{2^N}}}\right)^{-1}.$$

Then the billiard flow is ergodic on Q.

See [77] for many other references about ergodicity of the billiard flow.

In [67], Katok proves that $S(L)$ grows subexponentially. D. Scheglov [91] has the best refinement on this result to date:

Theorem 5.16. *With respect to almost every irrational triangle T, the following estimate on $S(L)$ holds:*

$$\lim_{L\to\infty} \frac{S(L)}{\exp(L^{\varepsilon})} = 0 \quad \forall \varepsilon > 0.$$

6. POLYGONAL OUTER BILLIARDS

B. H. Neumann [86] introduced outer billiards in the late 1950*s* and then J. Moser [83, 84] popularized it in the 1970s as a toy model for planetary motion. Outer billiards is a game that is played on the outside of a billiard table. Given a compact convex set $K \subset \boldsymbol{R}^2$ and a point $x_0 \in \boldsymbol{R}^2 - K$, one defines x_1 to be the point such that the segment $\overline{x_0 x_1}$ is tangent to K at its midpoint and K lies to the right of the ray $\overrightarrow{x_0 x_1}$. See Figure 6 for an example.

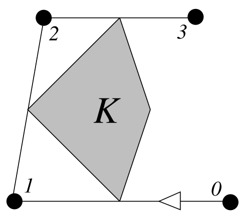

FIGURE 6
Polygonal outer billiards.

The iteration $x_0 \to x_1 \to x_2 \to \cdots$ is called the *forward outer billiards orbit* of x_0. The backward orbit is defined similarly. When K is a polygon, this map is well defined in the complement of finitely many rays which extend the sides of K.

6.1. Periodic orbits

When K is a polygon, the second iterate ψ of the outer billiards map is a piecewise translation. The translation vectors all have the form $2(v_i - v_j)$, where v_i and v_j are vertices. Every finite power of ψ is defined in the complement of finitely many lines. In particular, if ψ has a periodic orbit, then there is a maximal open convex set containing this point, often called a *periodic island*, which consists entirely of periodic points of the same period. This is somewhat akin to the phenomenon that periodic billiard paths on polygons come in infinite parallel families.

Periodic orbits are easier to come by in polygonal outer billiards. C. Culter proved the following pretty easy result. See [104].

Theorem 6.1. *Outer billiards with respect to any convex polygon has infinitely many periodic islands.*

Three teams of authors, namely Vivaldi–Shaidenko [110], Kolodziej [70], and Gutkin–Simanyi [55] proved the following result:

Theorem 6.2. *If K is a convex polygon with rational vertices then all outer billiards orbits on K are bounded. Hence all orbits are periodic.*

Proof. (Sketch of both results.) Let P be a convex polygon. Scale so that P has integer vertices. For simplicity, assume that P has no parallel sides. For each edge e of P, consider the strip Σ_e with the following description. One side of Σ_e contains e, and the unique vertex of P farthest from e lies on the centerline of the strip. This Σ_e is twice as wide as P in some sense. Let $e_1, \ldots, e_n$ be the edges of P ordered according to their slopes. Let $\Sigma_1, \ldots, \Sigma_n$ be the corresponding strips.

There are partitions of each strip Σ_j into parallelograms translation equivalent to $\Sigma_j \cap \Sigma_{j+1}$ such that, for p far away from P, some power of ψ maps each parallelogram in Σ_j isometrically into Σ_{j+1}. Figure 7 shows this.

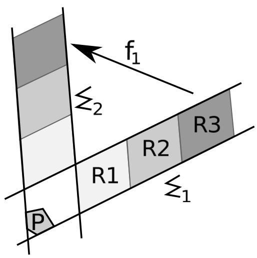

FIGURE 7
The strip map.

Let f_j be the map which maps the nth parallelogram in Σ_j to the nth parallelogram in Σ_{j+1}, as indicated by Figure 7. Let Ψ denote the first return map to the strip Σ_1. Outside a large compact set, Ψ agrees with $f_n \circ \cdots \circ f_1$. In particular, Ψ (and hence ψ) has a periodic orbit provided that, far away from P, there are parallelograms $R_{i_j} \subset \Sigma_j$ such that

$$H(i_1, \ldots, i_n) = R_1 \cap f_1^{-1}(R_2) \cap \cdots \cap f_1^{-1} \circ \cdots \circ f_n^{-1}(R_n) \neq \emptyset. \tag{6.1}$$

I like to think of these as "resonances."

If P has integer vertices, then for certain lists of integers $i_1, \ldots, i_n$, the set $H(i_1, \ldots, i_n)$ is convex set that completely spans Σ_1, like a tennis ball in a can, so that $\Sigma_1 - H$ is disconnected. The corresponding periodic island separates P from infinity, like a necklace. These special integer lists recur periodically, so there is an infinite sequence of these necklace barriers marching out to ∞. The other orbits are trapped between these barriers. Hence all orbits are bounded. But since ψ is locally a translation by integer vectors, all orbits are periodic by the pidgeonhole principle.

When P is arbitrary, these exact resonances do not occur, but infinitely often they nearly occur. Hence there are infinitely many periodic islands which very nearly span the strip Σ_1. ■

Question 6.3. Does outer billiards have a dense set of periodic islands with respect to almost every polygon?

6.2. Aperiodic orbits

Let us examine the proof sketch just given more carefully. The existence of the resonances producing the necklace orbits just discussed does not really depend on the polygon P having integer vertices. The important thing is that the consecutive parallelograms $\Sigma_j \cap \Sigma_{j+1}$ have commensurable areas. A polygon which has this property is called *quasi-rational*. Thus, the stronger version of Theorem 6.2 is that every quasirational polygon has all orbits bounded.

The regular polygons certainly are quasirational. Hence all the outer billiards orbits are bounded for regular polygons. In [102], S. Tabachnikov proved the following result:

Theorem 6.4. *Outer billiards orbits on the regular pentagon has some aperiodic orbits.*

Figure 8 shows the pattern of periodic islands for the regular pentagon. The outer 5 periodic islands, not entirely shown, form the first necklace. Theorem 6.4 is proved by establishing that the first return map to a certain triangular region T in the plane is a *renormalizable* polygon exchange map. In this case, this means that the first return map to some smaller triangle $T' \subset T$ is conjugate, via a similarity, to the first return map to T.

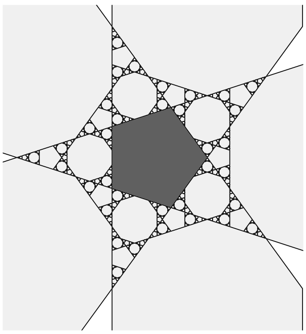

FIGURE 8
Periodic islands for the regular pentagon.

The cases $n = 8, 10, 12$ also have a self-similar structure. Without having a reference, I have the sense that the case $n = 7$ is somewhat understood in the sense that there are some regions of renormalization. I think that the cases $n = 9, 11$ are not understood at all. G. Hughes [63] has made beautiful and detailed pictures of outer billiards on regular polygons. These pictures (and earlier ones) suggest

Conjecture 6.5. *Outer billiards on the regular n-gon has an aperiodic orbit if $n \neq 3, 4, 6$.*

I think that this is not known aside from $n = 5, 8, 10, 12$, and perhaps $n = 7$.

D. Genin [50] made a thorough study of outer billiards on trapezoids, and found examples of open subsets of aperiodic orbits.

6.3. Unbounded orbits

One central problem in the subject is the *Moser–Neumann Problem:* Do there exist any outer billiards systems with unbounded orbits? In [84] and [83], Moser discussed this problem in terms of the stability of a toy problem for planetary motion.

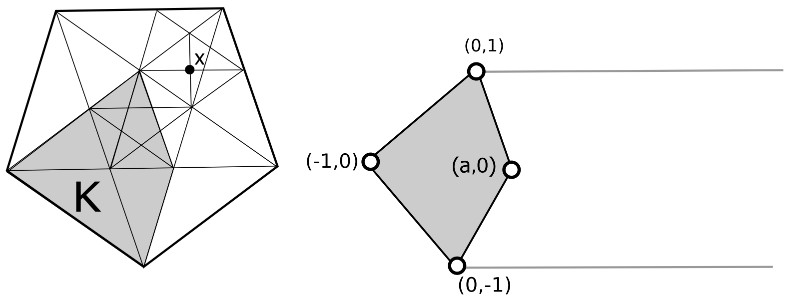

FIGURE 9
The Penrose kite (left) and the kite K_a (right).

In [93] I answered this question by showing that outer billiards with respect to the Penrose kite has an unbounded orbit. The left side of Figure 9 shows the Penrose kite and a point x with an unbounded orbit. The auxiliary lines are just scaffolding to show the construction.

Later, I proved a more general theorem in [95] which I will now describe. A *kite* is a convex quadrilateral with a line of symmetry that is a diagonal. The other diagonal divides K into two triangles, and the kite is *irrational* if these areas are irrationally related. Call an outer billiards orbit with respect to K *erratic* if it exits every compact subset of the plane and enters every open neighborhood of K.

Theorem 6.6. *There exist erratic orbits with respect to any irrational kite.*

Proof. (Very rough sketch.) Outer billiards is affinely natural, so it suffices to consider the kite K_a shown on the right side of Figure 9. Let Λ denote the two rays $[0, \infty) \times \{-1, 1\}$. Let Ψ denote the first return map to Λ. It suffices to prove that Ψ has unbounded orbits. Much like the continued fraction approximation, there is a canonical sequence of odd/odd rationals

$\{p_n/q_n\} \to a$ such that

$$|p_n q_{n+1} - p_{n+1} q_n| = 2. \tag{6.2}$$

Let $K_n = K_{p_n/q_n}$. Let Ψ_n be the corresponding first return map to Λ.

We partition Λ into intervals of length $2/q_n$ having centers $(1/q_n, \pm 1)$, $(3/q_n, \pm 1)$, etc. The map Ψ_n permutes these infinitely many intervals. We encode the combinatorial structure of this permutation as follows. There is a map from $\boldsymbol{Z}^2$ to our intervals defined as follows:

$$\Phi(m,n) = \left(\frac{2mp_n}{q_n} + 2n, (-1)^{m+n+1}\right).$$

I discovered that for each Ψ_n orbit there is an embedded nearest neighbor path on $\boldsymbol{Z}^2$ such that Φ maps consecutive vertices of the path to consecutive points of the orbit. I call this path the *arithmetic graph* of the orbit.

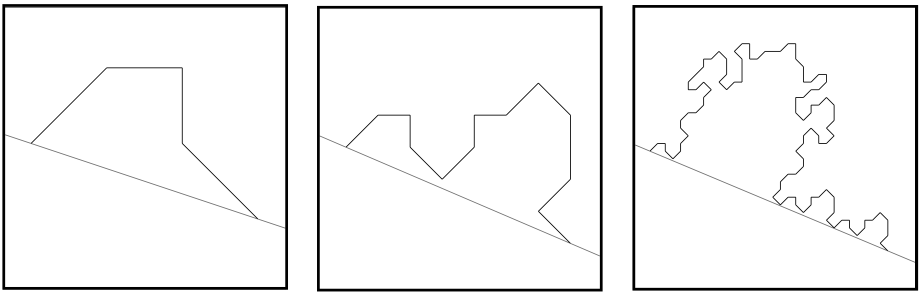

FIGURE 10
The arithmetic graphs $\Gamma(1/3)$ and $\Gamma(3/7)$ and $\Gamma(13/31)$.

Let $\Gamma_n = \Gamma(p_n/q_n)$ be the arithmetic graph of the orbit of $(1/q_n, 1)$. It is useful to think of Γ_n as a biinfinite path. One period of Γ_n connects $(0,0)$ to $(q_n, -p_n)$. (This statement requires p_n, q_n both to be odd.) The distance that Γ_n rises up above the line L_n of slope $-p_n/q_n$ through $(0,0)$ is comparable to the excursion distance of the corresponding Ψ_n orbit. Figure 10 shows one period of three of these graphs. Each pair of rationals satisfies equation (6.2). Notice that each graph copies at least one period of the previous one. Let us call this property *coherence*.

There are two main steps in the proof. The first one is to establish the coherence. The second step involves showing that the graph Γ_n rises up at least $q_n/2$ above the line L_n. Once we have these properties, we can take a limit as $n \to \infty$ and get an unbounded orbit. The fact that these graphs copy each other makes them oscillate away and back to L_n on the order of n times, with some of the oscillations being very large. This is the mechanism for the erratic orbits.

For each parameter $b \in (0,1)$ there is a 3-dimensional polyhedron exchange transformation $(\widehat{\Lambda}_b, \widehat{\Psi}_b)$ and a locally affine map $\Upsilon_b : \Lambda \to \widehat{\Lambda}_b$ such that

$$\widehat{\Psi}_b \circ \Upsilon_b = \Upsilon_b \circ \Psi.$$

In other words, Υ_b is a semiconjugacy between a 1-dimensional noncompact dynamical system and a 3-dimensional compact dynamical system. For almost every b, the image of Υ_b is dense. Thus, the 3-dimensional system is typically a kind of a dynamical compactification of the 1-dimensional system. Each dynamical system $(\widehat{\Lambda}_b, \Psi_b)$ in turn sits as a slice of a 4-dimensional integral affine polytope exchange transformation.

Step 1: For two parameters $b = p_n/q_n$ and $b' = p_{n+1}/q_{n+1}$ satisfying equation (6.2), the corresponding polyhedron exchange maps and semiconjugacies are close in the sense needed for the coherence phenomenon.

Step 2: The polyhedron exchange map $(\widehat{\Lambda}_{p_n/q_n}, \widehat{\Psi}_{p_n/q_n})$ is determined by some partition of $\widehat{\Lambda}_{p_n/q_n}$ into smaller polyhedra, and the walls of this partition give rise to two infinite grids H_n and H_n' in $\boldsymbol{R}^2$. It turns out that Γ_n can only cross lines of H_n near points where H_n and H_n' intersect. In particular, some line from H_n separates the endpoints of (one period of) Γ_n, and the only point of this line on H_n' is at least $q_n/2$ from the line L_n. ■

It turns out that the grid phenomenon in Step 2 above was just the tip of the iceberg. I eventually found a kind of combinatorial model for the arithmetic graphs discussed in the proof above. See [96].

Question 6.7. Does outer billiards have unbounded orbits with respect to almost every polygon?

6.4. Nonpolygonal domains

Let me say a few words about nonpolygonal outer billiards. D. Dolgopyat and B. Fayad [34] prove the following result:

Theorem 6.8. *Outer billiards has unbounded orbits with respect to the domain obtained by cutting a disk in half.*

There is some stability to the argument. Dolgopyat and Fayad more generally prove their result for domains obtained by nearly cutting a disk in half. The unbounded orbits look much different from my erratic orbits: there is just an open set of them which escapes straight to infinity. Kites and (near) semidisks are (up to affine transformations) the only known examples of shapes with respect to which outer billiards has unbounded orbits.

Question 6.9. Is there a strictly convex domain or a C^1-domain with respect to which outer billiards has unbounded orbits?

Using KAM theory (in an argument outlined by J. Moser), R. Douady [35] proved the following result:

Theorem 6.10. *Outer billiards has all bounded orbits with respect to C^6-oval of positive curvature.*

In a different direction, P. Boyland [17] gives examples of C^1-domains which have orbits that enter every open neighborhood of the domain. These orbits crash into the domain, in an asymptotic sense. The domains are not C^2.

One other kind of domain I got curious about is a closed convex domain in the projective plane that is invariant with respect to a surface group of projective automorphisms. The interiors of such domains are universal covers of the so-called *convex projective surfaces*. Typically, such curves are C^1 with a Hölder-continuous derivative.

7. OVALS

I will start by describing billiards inside an ellipse and related topics. See [36] for a comprehensive reference. Following this, I will move on to more exotic kinds of tables.

7.1. Billiards in an ellipse

Let E_1 be a noncircular ellipse. There are 2 special billiard paths on E_1 having period 2 and then a third special one which goes through the foci of E in-between bounces. Every other orbit is tangent to a confocal conic section E_2, either an ellipse or a hyperbola, called a *caustic*. The *phase portrait* nicely organizes all the billiard paths. The *phase space* Φ is the open cylinder of pairs (p, ℓ) where $p \in E_1$ and ℓ is a line through p which is not tangent to E_1 at p. The *billiards map* carries (p_1, ℓ_1) to (p_2, ℓ_2), where $\{p_1, p_2\}$ are the two points of $l_1 \cap E_2$ and $\{l_1, l_2\}$ are the two lines making the same angle with (the tangent line to) E_1 at p_1.

The left-hand side of Figure 11 shows the phase portrait and indicates both the special and generic orbits by letters. The right-hand side shows what the corresponding billiard paths look like. Note that the billiards map preserves the curves corresponding to orbits with ellipse caustic and the square of the billiards map preserves the curves corresponding to orbits with hyperbola caustic.

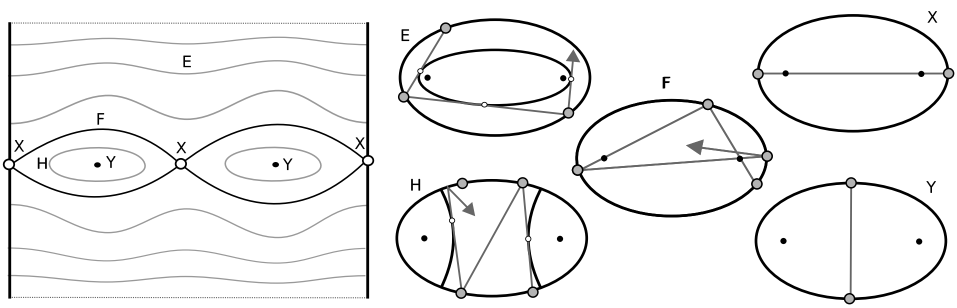

FIGURE 11
The phase portrait for elliptical billiards.

One well-known fact is that periodic billiard paths in an ellipse having the same caustic have the same perimeter. Experimenting with the computer recently, Dan Reznik has discovering an avalanche of related results. For instance, within a family of periodic billiard paths corresponding to the same caustic, the product of the cosines of the interior angles is constant. See [89] for an exposition of some of these results.

7.2. Poncelet and Cayley

The next result is a special case of the Poncelet Porism, which (when suitably phrased to avoid mentioning billiards) works for any pair of conics, confocal or not.

Theorem 7.1. *Let Ω be one of the smooth curves of the billiards phase space. If Ω corresponds to an ellipse caustic then the restriction of f to Ω is a rotation in suitable coordinates.*

Proof. (Sketch.) Let $E_1(\boldsymbol{C})$ and $E_2(\boldsymbol{C})$ denote the spheres which are extensions of E_1 and E_2 to the complex projective plane. Let $\Omega(\boldsymbol{C})$ denote the set of pairs (p, ℓ) where $p \in E_1(\boldsymbol{C})$ and ℓ is a complex line through p tangent to $E_2(\boldsymbol{C})$. The map $(p, \ell) \to p$ is a 2-fold holomorphic branched cover, branched over the 4 intersection points of the two spheres. Like all complex tori, $\Omega(\boldsymbol{C})$ is biholomorphic to a flat torus. The billiards map is an isometric rotation in these coordinates because it is the product of 2 holomorphic involutions, $(p, \ell)) \to (p, \ell')$ and $(p, \ell)) \to (p', \ell)$. Here ℓ' is the other line through p tangent to $E_2(\boldsymbol{C})$ and p' is the other point where ℓ intersects $E_1(\boldsymbol{C})$. See [53] for more details. ■

There is another approach to Theorem 7.1 which works specifically in the case when E_1, E_2 are confocal. Let E_1^* be the region bounded by E_1. There is a uniformizing change of coordinates [36, 38] somewhat akin to the Schwarz–Christoffel transform, which carries the relevant component of $E_1^* - E_2$ to either a flat cylinder or a rectangle. In these coordinates, the billiard paths with caustic E_2 transform to ordinary billiards which move parallel to the directions $(\pm 1, \pm 1)$. Changing the caustic E_2 changes the rectangle/cylinder.

Which caustics give rise to periodic billiard paths? Cayley's amazing answer works for any pair (E_1, E_2) of conics, confocal or not. Say that (E_1, E_2) *supports a Poncelet n-gon* if there exists a closed n-gon whose vertices are in E_1 and whose edges are contained in lines tangent to E_2. In homogeneous coordinates, E_k is the zero-set of an equation $\sum d_{ij} x_i x_j = 0$ encoded by a 3×3 matrix $D_k = \{d_{ij}\}$. Take the Taylor series expansion

$$\sqrt{\det(t D_1 + D_2)} = A_0 + A_1 t + A_2 t^2 + \cdots . \tag{7.1}$$

Theorem 7.2. *Let Δ_n to be the left (respectively right) determinant when $n = 2m + 1$ (respectively $n = 2m$)*

$$\begin{vmatrix} A_2 & \dots & A_{m+1} \\ \vdots & \ddots & \vdots \\ A_{m+1} & \cdots & A_{2m} \end{vmatrix}, \quad \begin{vmatrix} A_3 & \dots & A_{m+1} \\ \vdots & \ddots & \vdots \\ A_{m+1} & \cdots & A_{2m} \end{vmatrix}. \tag{7.2}$$

Then (E_1, E_2) supports a Poncelet n-gon if and only if $\Delta_n = 0$.

See [54] for a modern proof of Cayley's Theorem. In [39], V. Dragović and M. Radnović give the complete analogue of Theorem 7.2 for billiards in a higher-dimensional ellipsoid.

7.3. Piecewise elliptical tables

In [37, 38] Dragović and Radnović use the transformation mentioned in connection with the Poncelet Porism to study a more exotic situation in which the table is made from pieces of two confocal ellipses, as in Figure 12. They term this kind of billiards *pseudo-integrable*. (This term also refers to rational billiards in the physics literature.)

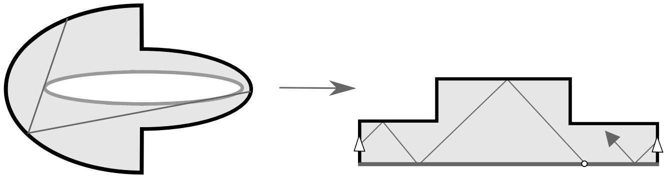

FIGURE 12
A pseudointegrable table.

Since the pieces are confocal, the billiard paths still have caustics. For each choice of caustic, the uniformizing map carries the domain which is either a right-angled polygon or a topological cylinder with a right-angled boundary. The billiard paths on these tables move parallel to $(\pm 1, \pm 1)$ as above. (My picture is just a cartoon; I did not compute the uniformizing map.) These systems exhibit a wider variety of behavior than integrable billiards [39], such as orbits which are dense but not equidistributed. See [48] and [47] for further developments.

7.4. The stadium

Figure 13 shows the *Bunimovich stadium*, another billiard table that has been intensely studied. This domain is convex but not strictly convex, and only C^1. The boundary of the stadium is a union of two semicircles and two line segments. This is really a 1-parameter family of examples. The parameter is the ratio of the line segment length to the semicircle length.

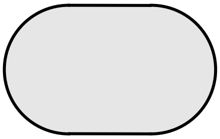

FIGURE 13
A Bunimovich stadium.

Here is a version of the theorem of Bunimovich which is easy to state:

Theorem 7.3. *Billiards in any stadium is ergodic. In particular, almost every billiard path is dense.*

This result is quite surprising because billiards in the disk is completely integrable. Once you introduce even the tiniest line segment, the billiard map changes completely. The full theorem of Bunimovich has more conclusions. See the paper of Misiurewicz and Zhang [82] for recent results about stadium billiards, and many other references.

7.5. Periodic orbits

Now I will talk about billiards in a general oval. From now on, by an *oval* I mean an infinitely differentiable and strictly convex closed curve. Many authors care about the exact level of differentiability. I am going to sweep this under the rug; you should consult the original sources for the precise generality needed for the results.

When C is an oval, we can define the phase space just as in the ellipse case. There is always an invariant area form on the phase space. It is given locally by $\sin(\theta) d\theta ds$, where θ is the angle that the relevant line ℓ in the pair (p, ℓ) makes with C, and ds is arc length. The following theorem of Birkhoff [13] vitally uses the area-preserving nature of the billiard map on phase space.

Theorem 7.4. *If C is an oval then, for every $n > 1$ and every integer rotation number $|r| < n$, there are at least 2 periodic billiard paths in C having period n and rotation number r.*

Area-preserving maps are special cases of symplectic maps. Sometimes one can use symplectic geometry to get results about billiards which seem (to me) impossible to get in a different way. Let me discuss one striking result along these lines, due to Y. Ostrover and S. Artstein-Avidan [2]. Let $\xi(K)$ denote the length of the shortest periodic billiard path in K. Given two sets K_1, K_2, define $K_1 + K_2$ to be the set of sums $v_1 + v_2$ with $v_1 \in K_1$ and $v_2 \in K_2$.

Theorem 7.5. *For any ovals K_1, K_2, we have $\xi(K_1 + K_2) \geq \xi(K_1) + \xi(K_2)$.*

The result in [2] is stated and proved for smooth convex domains in all dimensions.

7.6. Two guiding conjectures

Here I will discuss two geometric conjectures about billiards in convex ovals. Motivated by his theorem about the asymptotics of the eigenvalues of the Laplacian in a convex domain, V. Ya. Ivrii [64] made the following conjecture:

Conjecture 7.6. *Almost every billiard path in an oval is aperiodic.*

Ivrii's conjecture is wide open, but here is some partial progress. Y. Baryshnikov and V. Zharniksky [9] prove that there cannot exist an open set of 3-periodic orbits for an oval. M. Rychlik [90] proves the following result with the assistance of the computer:

Theorem 7.7. *The set of 3-periodic billiard paths in any oval has measure 0.*

L. Stojanov [101] removes the computer dependence, then M. Wojtkowski [113] and Ya. B. Vorobets [111] give different and independent proofs for more general domains.

In [52], A. Glutsyuk and Y. Kudryashov prove the following result:

Theorem 7.8. *No oval has an open set of* 4*-periodic billiard paths.*

In [71], V. F. Lazutkin proves that for any strictly convex and sufficiently smooth oval, there is a positive Lebesgue measure union of caustics for billiard paths in the oval. However, there are generally gaps between the caustics. The Birkhoff–Poritsky Conjecture is a rigidity conjecture which says essentially that if there are no gaps between the caustics then the table is an ellipse. Let Φ denote the phase space for billiards on the oval C.

Conjecture 7.9. *Let C be an oval. Suppose that some neighborhood of $\partial\Phi$ is foliated by invariant curves. Then C is an ellipse.*

The first progress is due to M. Bialy [10]:

Theorem 7.10. *If Φ is completely foliated by invariant curves then C is a circle.*

Say that an invariant curve in Φ is a q-curve if every orbit in the curve has period q. Recently, M. Bialy and A. Mironov [12] proved the following result:

Theorem 7.11. *Suppose C is centrally symmetric and there is a neighborhood N of $\partial\Phi$, foliated by invariant curves, such that ∂N is a union of two* 4*-curves. Then C is an ellipse.*

The smaller the neighborhood of the boundary, the higher the period of the orbit, so the neighborhood needed for this theorem is sort of medium-sized. One really neat fact proved along the way is that, given the hypotheses of the theorem, the billiard paths corresponding to points on the 4-curves are all parallelograms.

In another direction, A. Glutsyuk [51], extending work in Bialy–Mironov [11], has given a solution to the conjecture in a restricted case where the objects are not just smooth but algebraic. See [103] for a related result in the outer billiards case.

In [66], V. Kaloshin and A. Sorrentino have proved a completely general version of the Birkhoff Conjecture for ovals which are perturbations of ellipses. The main result in [66] pays careful attention to the level of differentiability; here is a corollary.

Theorem 7.12. *Suppose that C is sufficiently close to an ellipse in the C^∞-sense and also Φ has a q-curve for each $q = 3, 4, 5, 6, \ldots$ Then C is an ellipse.*

Referring to the discussion about Lazutkin's theorem, this last result allows there to be gaps between the caustics, but it still supposes a very particular kind of structure to certain of them.

7.7. The pentagram rigidity conjecture

I cannot resist bringing up a question I have been curious about for 30 years. The question intertwines the Poncelet Porism and the so-called *deep diagonal pentagram maps*. To me it seems like a discrete variant of the Birkhoff–Poritsky Conjecture. I will state the conjecture for the pair $(3, 8)$ just for simplicity. Figure 14 shows two 8-gons O_1 and O_2 related by a construction involving the 3-diagonals of O_1. The right-hand side indicates how

O_2 might not be convex even if O_1 is convex. This operation is best defined in the projective plane. (Convexity still makes sense there.)

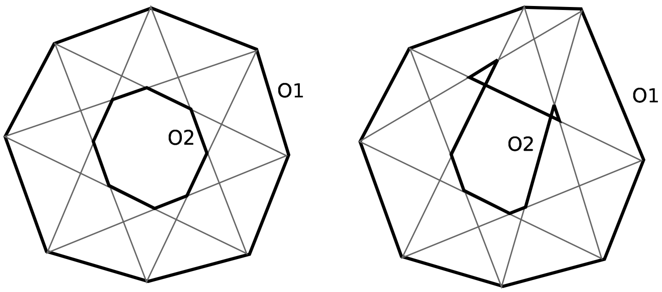

FIGURE 14
The 8-gons O_1 and O_2.

Starting with O_0 we can construct the biinfinite sequence $\{O_n\}$ of 8-gons, in which successive ones are related by the construction. If O_0 is a convex Poncelet polygon, then O_k is a projectively equivalent convex Poncelet polygon for all $k \in \boldsymbol{Z}$.

Conjecture 7.13. *If O_k is convex for all k then O_0 is a Poncelet polygon.*

A. Izosimov [65] has made a bit of general progress related to this conjecture by showing that if n is odd and two convex polygons related by the $(2, n)$ construction are projectively equivalent then they both are Poncelet polygons. I recently [97] established the very special case when the octagons have 4-fold rotational symmetry.

8. TABLES WITH OBSTACLES

The subject of dispersive billiards is an enormous one and this small chapter just gives you a taste. See [24] for a survey. One of the main themes in the subject is understanding the mixing properties of the system. Another main theme is the attempt to give rigorous mathematical foundations for physical processes like Brownian motion and the transfer of mass and heat in a gas.

8.1. Mixing

To build some intuition for dispersive billiards, consider how some given element $T \in \mathrm{SL}_2(\boldsymbol{Z})$, with an eigenvalue $\lambda > 1$, acts on the square torus $Y = \boldsymbol{R}^2/\boldsymbol{Z}^2$. Let us show that T is mixing in the sense given by the left-hand side of equation (3.4).

Lemma 8.1. *T is mixing.*

Proof. (Sketch). I will just consider the case when U is a parallelogram with sides parallel to the eigenvectors of T. Let $|U|$ be the side length of U. Corresponding to λ there is an irrational invariant geodesic foliation of Y. For large n, the set $T^n(U)$ is a long thin parallelogram smeared out along this foliation. The long side of $T^n(U)$ is about $|U|\lambda^n$ in length and the short side is about $|U|\lambda^{-n}$ in length. So, $T^n(U)$ is essentially a really thin strip that closely follows an irrational geodesic for a really long time. We have already seen that the irrational geodesics in Y are equidistributed. Given this property, $T^n(U)$ spends about $\mu(U)$ percent of the time in V. ■

Given the exponential growth of the length of $T^n(U)$, the quantity on the left-hand side of equation (3.4) decays exponentially in n. That is, at least when U and V are nice sets like rectangles or disks the quantity on the left side of equation (3.4) is of the order of e^{-Cn} for some $C > 0$. This kind of decay, suitably generalized and formalized, is called *exponential mixing*. See [88] for a definition. Mixing is stronger than ergodicity, and exponential mixing is even stronger than that.

8.2. The Lorentz gas

The classic *Lorentz gas*, also known as a *Sinai billiard* [99], is a billiard ball bouncing around on the table you get by removing a round disk D from the center of a square.

Theorem 8.2. *The billiard map on* $[0,1]^2 \setminus D$ *is mixing.*

Proof. (Very rough sketch.) Ignoring the measure zero set of billiard paths which avoid D, we can define the phase space Φ of the system to be the cylinder of pairs (p, ℓ) where $p \in \partial D$ and ℓ is a line through p not tangent to ∂D. The same measure as for smooth ovals is an invariant one for the system.

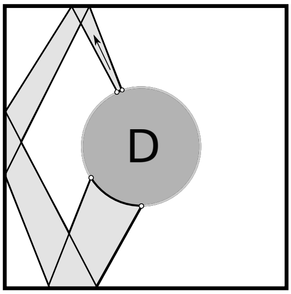

FIGURE 15
Scattering property in action.

Each billiard path that leaves ∂D bounces some number of times on the square and then returns to ∂D. Some word records the intermediate bounces. Partition Φ by the elements that correspond to the same word. Consider an arc of elements of the same partition which leave ∂D at the same angle, as shown in Figure 15. These elements spread out before returning to ∂D. Since the billiards map is area preserving, it stretches each partition piece in one direction and compresses it in the other. The longer the word, the more dramatic the effect. So, the local behavior is like that considered for the map T considered in Lemma 8.1. ■

More generally, you could remove finitely many disjoint strictly convex scatterers from the interior of the unit square or from a flat torus. The table has *finite horizon* if all billiard paths hit the scatterers. The mixing properties of billiards on these tables—or at least when the properties were established—depend on the finite horizon property and also on whether one considers the billiards map or the billiards *flow* on the unit tangent bundle. Here is a rundown of the results:

(1) Y. Young [117] shows that the billiard map is exponentially mixing in the finite horizon case, and then N. Chernov [23] establishes this in the infinite horizon case.

(2) V. Baladi, M. Demers, and C. Liverani [7] establish the exponential mixing for the flow in the finite horizon case and P. Bálint, O. Butterley, and I. Melbourne [8] establish polynomial mixing in the infinite horizon case.

A more complicated situation arises when the scatterers are allowed to touch. J. de Simoi and P. Toth [32] prove that the billiards map is exponentially mixing in the finite horizon case when no scatterers are tangent. In [25], N. Chernov and H.-K. Zhang show that the billiard map is polynomially mixing in the finite horizon case when tangencies are allowed.

Here are some poorly understood situations in this area. One thing you can do is play billiards in the plane, after removing an infinite number of scatterers but not in a periodic pattern. (The periodic case is the universal cover of the kind of the example considered above.) Another thing you can do is replace a single bouncing point with several or many bouncing disks of finite size. D. Dolgopyat and P. Nándori [33] make some recent progress in the case of 2 disks.

8.3. Breakout

Let me close this survey with some whimsical questions. Inspired by the video game *Breakout* [18,87], one could imagine a ball bouncing around an infinite periodic array of disk scatterers but with the twist that a scatterer disappears as soon as it is hit.

Question 8.3. Does a typical billiard path erase all the scatters eventually?

Here is one thing I noticed about the breakout game when it is played on the 1-skeleton of the infinite square tiling. (Again, a reflector disappears as soon as it is hit.)

Conjecture 8.4. *If you start the ball moving with slope* $\sqrt{2}$*, the billiard path eventually erases all the reflectors.*

These systems remind me a little bit of Langton's ant, and the questions about them seem to verge on the territory of cellular automata.

ACKNOWLEDGMENTS

Many people helped me write this survey. I would like to think Jayadev Athreya, Misha Bialy, Dan Cristofaro-Gardiner, Mark Demers, Vladimir Dragovic, Giovanni Forni, Pat Hooper, Howard Masur, Curtis McMullen, Péter Nándori, Dan Reznik, Sergei Tabachnikov, and Barak Weiss for very helpful discussions. I listened to all these folks but I did not always follow their advice; any errors in the exposition and omissions are my fault.

FUNDING

This work is supported by the U. S. National Science Foundation grant DMS-2102802.

REFERENCES

[1] P. Apisa and A. Wright, Marked points on translation surfaces. 2021, arXiv:1708.03411.

[2] S. Artstein-Avidan and Y. Ostrover, Bounds for Minkowski billiard trajectories in convex bodies. 2012, arXiv:1111.2353.

[3] J. S. Athreya, D. Aulicino, W. Hooper (appendix by A. Randecker), Platonic solids and high genus covers of lattice surfaces. *Exp. Math.* (2018), 1–51.

[4] A. Avila and V. Delacroix, Weak mixing directions in non-arithmetic Veech surfaces. *J. Amer. Math. Soc.* **29** (2016).

[5] A. Avila and G. Forni, Weak mixing for interval exchange transformations and translation flows. *Ann. of Math.* **165** (2007), 637–664.

[6] M. Bainbridge, Euler characteristics of Teichmüller curves in genus two. *Geom. Topol.* **11** (2007), 1887–2073.

[7] V. Baladi, M. Demers, and C. Liverani, Exponential decay of correlations of finite horizon Sinai billiard flows. *Invent. Math.* 2018.

[8] P. Bálint, O. Butterley, and I. Melbourne, Polynomial decay of correlations for flows, including Lorentz gas examples. *Comm. Math. Phys.* **368** (2019), 55–111.

[9] Y. Baryshnikov and V. Zharniksky, Billiards and nonholonomic distributions. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **300** (2004), 56–64.

[10] M. Bialy, Convex billiards and a theorem of E. Hopf. *Math. Z.* **214** (1993), 147–154.

[11] M. Bialy and A. Mironov, Angular billiard and algebraic Birkhoff conjecture. *Adv. Math.* **313** (2017), 102–126.

[12] M. Bialy and A. Mironov, The Birkhoff–Poritsky conjecture for centrally symmetric billiard tables. 2020, arXiv:2008.03566.

[13] G. D. Birkhoff, On the periodic motions of dynamical systems. *Acta Math.* (1927).

[14] M. Boshernitzan, G. Galperin, T. Krüger, and S. Troubetzkoy, Periodic billiard orbits are dense in rational polygons. *Trans. Amer. Math. Soc.* **350** (1998), no. 9, 3523–3535.

[15] I. Bouw and M. Möller, Teichmüller curves, triangle groups, and Lyapunov exponents. *Ann. of Math.* (2010).

[16] J. P. Bowman, Teichmüller geodesics, Delaunay triangulations, and Veech groups, Teichmüller theory and moduli problems. *Ramanujan Math. Soc. Lect. Notes Ser.* **10** (2020), 113–129.
[17] P. Boyland, Dual billiards, twist maps, and impact oscillators. *Nonlinearity* **9** (1996), 1411–1438.
[18] X. Bressaud and M.-C. Fournier, Casse-Briques. *Exp. Math.* (2015).
[19] A. Calderon, S. Coles, D. Davis, J. Lanier, and A. Oliveira, How to hear the shape of a billiard table. 2018, arXiv:1806.09644.
[20] C. Calta, Veech surfaces and complete periodicity in genus 2. *J. Amer. Math. Soc.* **17** (2004), 871–908.
[21] J. Chaika and G. Forni, Weakly mixing polygonal billiards. 2020, arXiv:2003.00890.
[22] J. Chaika, J. Smillie, and B. Weiss, Tremors and horocycle dynamics on the moduli space of translation surfaces. 2020, arXiv:2004.04027.
[23] N. Chernov, Decay of correlations in dispersing billiards. *J. Stat. Phys.* **94** (1999), 513–556.
[24] N. Chernov and R. Markarian, Chaotic billiards. *Math. Surveys Monogr.* **127**, 2006.
[25] N. Chernov and H.-K. Zhang, Improved estimates for correlations in billiards. *Comm. Math. Phys.* **277** (2008), 305–321.
[26] B. Cipra, R. Hanson, and A. Kolan, Periodic trajectories in right-triangle billiards. *Phys. Rev. E* **52** (1995), no. 2.
[27] D. Davis, Cutting sequences, regular polygons, and the Veech group. *Geom. Dedicata* **162** (2013), no. 1, 231–261.
[28] D. Davis and S. Lelièvre, Periodic paths on the pentagon, double pentagon, and golden L. 2018, arXiv:1810.11310.
[29] D. Davis, I. Pasquinelli, and C. Ulcigrai, Cutting sequences on Bouw–Möller surfaces: an S-adic characterization. *Ann. Sci. Éc. Norm. Supér.* **52** (2019), no. 4, 927–1023.
[30] V. Delacroix and W. P. Hooper, Flat surfaces in SageMath github, 2020, https://github.com/videlec/sage-flatsurf.
[31] P. Deligne and G. D. Mostow, Commensurabilities among lattices in $\mathrm{PU}(1,n)$. *Ann. of Math. Stud.* **132** (1993).
[32] J. De Simoi and I. P. Toth, An expansion estimate for dispersing planar billiards with corner points. *Ann. Henri Poincaré* **15** (2014), 1223–1243.
[33] D. Dolgopyat and P. Nándori, The first encounter of two billiard particles of small radius. 2016, arXiv:1603.07590.
[34] D. Dolyopyat and B. Fayad, Unbounded orbits for semicircular outer billiards. *Ann. Henri Poincaré* **10** (2009), 357–375.
[35] R. Douady, *These de 3-eme cycle*, Université de Paris 7, 1982.
[36] V. Dragović and M. Radnović, *Poncelet porisms and beyond: integrable billiards, hyperelliptic Jacobians, and pencils of quadrics*. Springer, 2011.

[37] V. Dragović and M. Radnović, Pseudo-integrable billiards and arithmetic dynamics. *J. Mod. Dyn.* **8** (2014), no. 1, 109–132.

[38] V. Dragović and M. Radnović, Periods of pseudo-integrable billiards. *Arnold Math. J.* **1** (2015), no. 1, 69–73.

[39] V. Dragović and M. Radnović, Periodic ellipsoidal polynomials and extremal polynomials. *Comm. Math. Phys.* **372** (2019), no. 1, 183–211.

[40] M. Duchin, V. Erlandsson, C. J. Leininger, and C. Sadanand, You can hear the shape of a billiard table: symbolic dynamics and rigidity for flat surfaces. 2019, arXiv:1804.05690.

[41] B. Edwards, S. Sanderson, and T. A. Schmidt, Computing Veech groups. 2021, arXiv:2012.12444.

[42] A. Eskin and H. Masur, Asymptotic formulas on flat surfaces. *Ergodic Theory Dynam. Systems* **21** (2001), no. 02, 443–478.

[43] A. Eskin and M. Mirzakhani, Invariant and stationary measures for the $SL_2(\boldsymbol{R})$ action on moduli space. *Publ. Math. Inst. Hautes Études Sci.* **127** (2018), 95–324.

[44] A. Eskin, M. Mirzakhani, and A. Mohammadi, Isolation, equidistribution, and orbit closures for the $SL_2(\boldsymbol{R})$ action on moduli space. *Ann. of Math.* **182** (2015).

[45] S. Filip, Splitting mixed Hodge structures over affine invariant manifolds. *Ann. of Math.* **183** (2016), no. 2, 681–713.

[46] G. Forni and C. Matheus, Introduction to Teichmüller theory and its applications to dynamics of interval exchange transformations, flows of surfaces, and billiards. 2013, arXiv:1311.2758.

[47] K. Fraczek, Recurrence for smooth curves in the moduli space and application to the billiard flow on nibbled ellipses. *Anal. PDE* **14** (2021), no. 4, 793–821.

[48] K. Fraczek, R. Shi, and C. Ulcigrai, Genericity on curves and applications: pseudo-integrable billiards, Eaton lenses and gap distributions. *J. Mod. Dyn.* **12** (2018).

[49] G. A. Galperin, A. M. Stepin, and Y. B. Vorobets, Periodic billiard trajectories in polygons. *Russian Math. Surveys* **47** (1001), 5–80.

[50] D. Genin, *Regular and chaotic dynamics of outer billiards*. Ph.D. thesis, Pennsylvania State University, State College, 2005.

[51] A. Glutsyuk, On polynomially integrable Birkhoff billiards on surfaces of constant curvature. 2019, arXiv:1706.04030.

[52] A. Glutsyuk and Y. Kudryashov, No planar billiard possesses an open set of quadrilateral trajectories. *J. Mod. Dyn.* **6** (2012), no. 3.

[53] P. Griffiths and J. Harris, Poncelet theorem in space. *Comment. Math. Helv.* **52** (1977), no. 2, 145–160.

[54] P. Griffiths and J. Harris, On Cayley's explicit solution to the Poncelet porism L'Enseignement. *Enseign. Math. (2)* **24** (1978), 1–2, 31–40.

[55] E. Gutkin and N. Simanyi, Dual polygonal billiard and necklace dynamics. *Comm. Math. Phys.* **143** (1991), 431–450.

[56] R. K. Guy, In *Unsolved problems in number theory*, pp. 365–366, Springer, 2004.

[57] L. Halbeisen and N. Hungerbuhler, On periodic billiard trajectories in obtuse triangles. *SIAM Rev.* **42** (1986), no. 2, 657–670.
[58] W. P. Hooper, *On the stability of periodic billiard paths in triangles*. Ph.D. thesis, Stony Brook, 2006.
[59] W. P. Hooper, Lower bounds on growth rates of periodic billiard trajectories in some irrational triangles. *J. Mod. Dyn.* **1** (2007), no. 4, 649–663.
[60] W. P. Hooper, Periodic billiard paths in right triangles are unstable. *Geom. Dedicata* **125** (2007), 39–46.
[61] W. P. Hooper, Grid graphs and lattice surfaces. *Int. Math. Res. Not.* **2013** (2013), 2657–2698.
[62] W. Hooper and R. E. Schwartz, Billiards in nearly isosceles triangles. *J. Mod. Dyn.* **3** (2009), no. 2, 159–231.
[63] G. Hughes, Dynamics of polygons. http://www.dynamicsofpolygons.org.
[64] V. Ya. Ivrii, The second term of the spectral asymptotics for a Laplace–Beltrami operator on manifolds with boundary. *Funct. Anal. Appl.* **14** (1980), no. 2, 98–106.
[65] A. Izosimov, The pentagram map, Poncelet polygons, and commuting difference operators. 2021, arXiv:1906.10749.
[66] V. Kaloshin and A. Sorrentino, On local Birkhoff conjecture for convex billiards. *Ann. of Math.* **188** (2018), 1–66.
[67] A. Katok, The growth rate for the number of singular and periodic orbits of a polygonal billiard. *Comm. Math. Phys.* **111** (1987), 151–160.
[68] A. Katok, Five more resistent problems in dynamics. http://www.personal.psu.edu/axk29/pub/5problems-expanded.pdf.
[69] S. Kerkhoff, H. Masur, and J. Smillie, Ergodicity of billiard flows and quadratic differentials. *Ann. of Math.* **124** (1986), 293–311.
[70] Kolodziej, The antibilliard outside a polygon. *Bull. Pol. Acad. Sci. Math.* **37** (1994), 163–168.
[71] V. F. Lazutkin, The existence of caustics for a billiard problem in a convex domain. *Izv. Akad. Nauk SSSR Ser. Mat.* **37** (1973).
[72] S. Lelièvre, T. Monteil, and B. Weiss, Everything is illuminated. *Geom. Topol.* **20** (2016), no. 3, 1737–1762.
[73] A. Leutbecher, Über die Heckeschen Gruen (λ) II. *Math. Ann.* **211** (1974), 63–86.
[74] H. Masur, Lower bounds for the number of saddle connections and closed trajectories of a quadratic differential. In *Holomorphic Functions and Moduli 1*, edited by D. Drasin, pp. 215–228, Springer, 1988.
[75] H. Masur, The Growth rate of trajectories of a quadratic differential. *Ergodic Theory Dynam. Systems* **10** (1990), 151–176.
[76] H. Masur, Hausdorff dimension of the set of nonergodic foliations of a quadratic differential. *Duke Math. J.* **66** (1992), 387–442.
[77] H. Masur and S. Tabachnikov, Rational billiards and flat structures. In *Handbook of Dynamical Systems 1, Part A, Chapter 13*, pp. 1015–1089, Elsevier, 2002.

[78] C. McMullen, Teichmüller curves in genus two: discriminant and spin. *Math. Ann.* **333** (2005), 87–130.

[79] C. McMullen, Teichmüller curves in genus two: torsion divisors and ratios of sines. *Invent. Math.* **165** (2006).

[80] C. McMullen, Billiards, heights, and the arithmetic of non-arithmetic groups. Preprint, 2020.

[81] C. McMullen, R. Mukamel, and A. Wright, Cubic curves and totally geodesic subvarieties of moduli space. *Ann. of Math.* **185** (2017), 957–990.

[82] M. Misiurewicz and H.-K. Zhang, Topological entropy of Bunimovich stadium billiards. 2020, arXiv:2006.03882.

[83] J. Moser, *Stable and random motions in dynamical systems, with special emphasis on celestial mechanics*. Ann. of Math. Stud. 77, Princeton University Press, Princeton, NJ, 1973.

[84] J. Moser, Is the solar system stable? *Math. Intelligencer* **1** (1978), 65–71.

[85] R. E. Mukamel, Orbifold points on Teichmüller curves and Jacobians with complex multiplication. *Geom. Topol.* **18** (2014), no. 2, 779–829.

[86] B. H. Neumann, *Sharing ham and eggs*, Summary of a Manchester Mathematics Colloquium, 25 Jan 1959, published in Iota, the Manchester University Mathematics. Students' Journal.

[87] E. Newkirk, Billiards with bombs. 2015, arXiv:1506.00683.

[88] M. Pollicott, Exponential mixing for the geodesic flow on hyperbolic three-manifolds. *J. Stat. Phys.* **67** (1992), 667–673.

[89] D. Reznik, R. Garcia, and J. Koillar, Can the elliptic billiard still surprise us? *Math. Intelligencer* **42** (2020), no. 1.

[90] M. R. Rychlik, Periodic points of the billiard ball map in a convex domain. *J. Differential Geom.* **30** (1989), 191–205.

[91] D. Scheglov, Complexity growth of a typical triangular billiard is weakly exponential. *J. Anal. Math.* **142** (2020), 105–124.

[92] R. E. Schwartz, Obtuse triangular billiards I: near the $(2,3,6)$ triangle. *Exp. Math.* (2006).

[93] R. E. Schwartz, Unbounded orbits for outer billiards. *J. Mod. Dyn.* **3** (2007).

[94] R. E. Schwartz, Obtuse triangular billiards II: 100 degrees of periodic billiard paths. *Exp. Math.* (2008).

[95] R. E. Schwartz, Outer billiards on kites. *Ann. of Math. Stud.* **171** (2009).

[96] R. E. Schwartz, The plaid model. *Ann. of Math. Stud.* **198** (2018).

[97] R. E. Schwartz, A textbook case of pentagram rigidity. 2021, arXiv:2108.07604.

[98] C. Series, The modular surface and continued fractions. *J. Lond. Math. Soc.* **31** (1985), no. 1.

[99] Ya. G. Sinai, Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards. *Russian Math. Surveys* **25** (1970), 137–189.

[100] J. Smillie and C. Ulcigrai, Beyond Sturmian sequences: coding linear trajectories in the regular octagon. *Proc. Lond. Math. Soc.* **101** (2020), no. 1.

[101] L. Stojanov, Note on the periodic points of the billiard. *J. Differential Geom.* **34** (1991).
[102] S. Tabachnikov, *Billiards*, Soc. Math. de Fr., "Panoramas et Syntheses" 1, 1995.
[103] S. Tabachnikov, On algebraically integrable outer billiards. 2007, arXiv:0708.0255.
[104] S. Tabachnikov, A proof of Culter's theorem on the existence of periodic orbits in polygonal outer billiards. 2007, arXiv:0706.1003.
[105] W. P. Thurston, *Shapes of polyhedra and triangulations of the sphere*. Geom. Topol. Monogr., 1998.
[106] G. Tokarsky, J. Garber, B. Marinov, and K. Moore, One hundred and twelve point three degree theorem. 2018, arXiv:1808.06667.
[107] S. Troubetzkoy, Periodic billiard orbits in right triangle II. 2005, arXiv:math/0505637.
[108] W. A. Veech, Teichmüller curves in moduli space, Eisenstein series, and an application to triangular billiards. *Invent. Math.* **97** (1989), 553–583.
[109] W. A. Veech, The billiard in a regular polygon. *Geom. Funct. Anal.* **2** (1992), no. 3.
[110] F. Vivaldi and A. Shaidenko, Global stability of a class of discontinuous dual billiards. *Comm. Math. Phys.* **110** (1987), 625–640.
[111] Ya. Vorobets, On the measure of the set of periodic points of a billiard. *Math. Notes* **55** (1994).
[112] Ya. Vorobets, Ergodicity of billiards in polygons. *Mat. Sb.* **188** (1997), no. 3, 117–171.
[113] M. P. Wojtkowski, Two applications of Jacobi fields to the billiard ball problem. *J. Differential Geom.* **40** (1994), 155–164.
[114] A. Wolecki, Illumination in rational billiards. 2019, arXiv:1905.09358.
[115] A. Wright, From rational billiards to dynamics on moduli spaces. *Bull. Amer. Math. Soc.* **53** (2016), no. 1, 41–56.
[116] A. Wright, Schwarz triangle mappings and Teichmüller curves: the Veech–Ward–Bouw–Möller curves. 2013, arXiv:1203.2685.
[117] Y. Young, Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math.* **147** (1998), 585–650.
[118] A. Zorich, Flat surfaces. In *Frontiers in number theory, physics, and geometry 1*, edited by P. Cartier, B. Julia, P. Moussa, and P. Vanhove, pp. 439–586, Springer, 2006.
[119] A. Zorich, The magic wand theorem of A. Eskin and M. Mirzakhani. *Gaz. Math.* **142** (2014), 39–54.

RICHARD EVAN SCHWARTZ

Department of Mathematics, Brown University, 151 Thayer Street, Box 1917, Providence, RI 02912, USA, Richard.Evan.Schwartz@gmail.com