What is quantum unique ergodicity?

Andrew Hassell[1]

Abstract

A (somewhat) nontechnical presentation of the topic of quantum unique ergodicity (QUE) is attempted. I define classical ergodicity and unique ergodicity, discuss their quantum analogues, and describe some recent breakthroughs in this young research field.

1. Introduction

The recently developed research field of quantum unique ergodicity, or QUE, has gained prominence due to the 2010 Fields Medal awarded to Elon Lindenstrauss, one of whose main achievements was the solution of the QUE question in a particular case. So what is QUE?

To describe this I will first try to describe *classical* unique ergodicity, and then explain the quantum version of this setup. First of all, ergodic theory is concerned with a measure space (X, μ), such that $\mu(X)$ is finite, together with a measurable transformation $T \colon X \to X$ which is measure preserving: the measure of $T^{-1}(A)$ is equal to the measure of A for every measurable subset A of X.

Example 1.1. A simple example is: X is the circle, that is, the interval $[0, 2\pi]$ with 0 and 2π identified, and T is the map $\theta \mapsto \theta + \alpha$ modulo 2π, that is, a rotation by angle α, for fixed $\alpha \in [0, 2\pi)$.

Example 1.2. A more interesting example is *billiards*: let B be a plane domain, that is, a compact set in $\mathbb{R}^2$ with smooth boundary. Actually we don't need the boundary to be infinitely smooth; it would be enough for a tangent line to exist at every point (or even almost every point) of the boundary. We define X to be the product of B with the circle S^1, and think of X as the set of unit tangent vectors to points in B. Similarly, μ is defined to be the product of Lebesgue measure on B and the standard measure on S^1. We will call μ the 'uniform measure'. Then for each real t, thought of as 'time', we define a map $\Phi_t \colon X \to X$, the 'billiard flow at time t', as follows: let $x \in B$ and $\theta \in S^1$. Then $\Phi_t(x, \theta)$ is the final position and direction of a billiard ball that starts at x with initial direction θ, and travels for time t in a straight line at unit speed in the interior of B, bouncing at the

[1]Department of Mathematics, Australian National University, Canberra ACT 0200.
Email: Andrew.Hassell@anu.edu.au

boundary and obeying the usual law of reflection (angle of reflection equals angle of incidence). See Figure 1. It turns out, though this is not obvious, that Φ_t is measure preserving for each t. This dynamical system (X, Φ_t) is called 'billiards on B'.
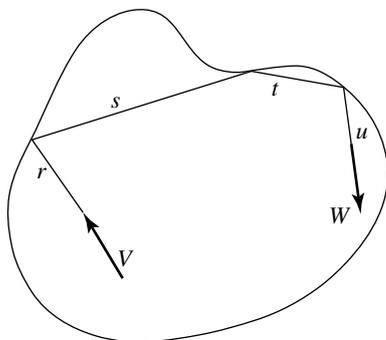

Figure 1. The billiard flow. The map ϕ_T maps the unit tangent vector V to the unit tangent vector W, where r, s, t, u denotes the lengths of the corresponding line segments and $T = r + s + t + u$.

Example 1.3. Example 1.2 has a geometric generalisation: we let X be the set of unit tangent vectors on a Riemannian manifold M, μ be the standard measure (cooked up out of the metric) on X, and Φ_t be geodesic flow.

Returning to the general setting, the transformation T is said to be ergodic if one of the following equivalent conditions holds.

- There is no decomposition of X into two disjoint pieces, both of positive measure, such that each is invariant under T.
- Almost every orbit is equidistributed: that is, for almost every x, we have

$$\lim_{N \to \infty} \frac{\#\{j \in \mathbb{N} \mid 0 \le j \le N - 1, \ T^j(x) \in A\}}{N} = \frac{\mu(A)}{\mu(X)} \qquad (1.1)$$

  for every measurable set $A \subset X$.
- Every T-invariant function is constant almost everywhere. (We say that a real function $f \colon X \to \mathbb{R}$ is T-invariant if $f \circ T = f$.)

Example 1.4. Consider billiards on a circular disc, and let T be the billiard flow Φ_1 at time 1. Consider the function $k \colon X \to \mathbb{R}$ that maps (x, θ) to the angle that the billiard orbit emanating from (x, θ) makes with the boundary (at any point where it meets the boundary); notice that this angle is the same at each such point. Then, this function is invariant under T, so we have found an invariant function k that isn't constant almost everywhere. Therefore billiards on a disc is not ergodic. In fact, it has exactly the opposite character: it is *completely integrable*, which means that the three-dimensional space X is foliated by 'invariant tori', that is, two-dimensional tori that are each invariant by the flow Φ_t, and therefore by T. In this case, each torus is just a level set of k.

On the other hand, here are two examples of ergodic billiards.

Example 1.5. The 'Barnett stadium' (as named by Peter Sarnak), introduced in [2], is bounded by two straight lines meeting at right angles, together with two circular arcs meeting the straight lines at right angles and meeting each other at a positive acute angle (see Figure 3). The domain has no symmetries. This billiard was proved by Sinai to be ergodic [13]. It is also Anosov, or strongly chaotic, meaning that nearby orbits tend to diverge at an exponential rate.

Example 1.6. The stadium is the plane domain consisting of a rectangle with two semicircles attached to two opposite sides of the rectangle. See Figure 2. The stadium billiard was proved by Bunimovich to be ergodic [3], and for this reason the domain is sometimes called the Bunimovich stadium. Notice that this domain certainly does *not* have the Anosov property that Example 1.5 has. In fact, there is a family of billiard trajectories that bounce vertically between the two parallel straight sides of the boundary; these are colloquially called *bouncing ball trajectories*. At these trajectories, nearby orbits only diverge at a linear rate, contrary to the Anosov condition.
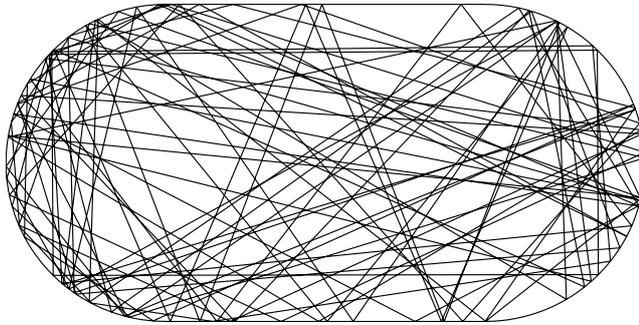

Figure 2. Billiards on a stadium, with an equidistributed orbit.

Our last 'classical' topic is to define classical unique ergodicity. In the remainder of this article, we shall assume some extra structure on our measure space (X, μ), namely, we assume that X is also a compact topological space, that μ is a finite Borel measure, and that all nonempty open sets of X have positive measure. Moreover, we shall assume that, as well as being measure-preserving, T is a homeomorphism, that is, continuous with continuous inverse. Notice that these assumptions are satisfied in all of our examples above.

Then, we say that (X, μ, T) is *uniquely ergodic* if μ is the only finite Borel measure on X invariant under T. Note that the word 'finite' is all-important here. Indeed, let $x \in X$ be fixed, and consider just atomic measures supported at points $T^n x$, for all integers n. For this measure to be invariant, each point must be given equal mass. So any orbit supports an invariant measure, but it is only finite if $T^n x = x$ for some $n \geq 1$, that is, if x is a periodic point. Thus we see that any invertible

transformation with a periodic point, (X, T, μ), is not uniquely ergodic. So for example, in Example 1.1, if $\alpha/(2\pi)$ is rational, then every point is periodic and then this example is not uniquely ergodic. On the other hand, if $\alpha/(2\pi)$ is irrational then it *is* uniquely ergodic — this follows from Weyl's equidistribution theorem (see [14, Chapter 4]). In general, unique ergodicity is a very strong property in dimensions two or higher. For example, neither Examples 1.5 nor 1.6 are uniquely ergodic, since both have periodic trajectories.

2. Quantum billiard systems

The word 'quantisation' has many meanings, but here we use the original meaning coming from quantum mechanics. In classical mechanics a state is specified by a point in phase space, giving both the position and momentum of a particle. If the particle is confined to a two-dimensional domain B (that is, a billiard), then a state consists of a point in B together with a tangent vector[a] at that point. On the other hand, the possible states of a quantum-mechanical 'particle' confined to B are given by complex-valued wave functions ψ on defined on B, with L^2 norm equal to 1, and the probability of finding the particle in a set contained in B is given by the integral of $|\psi|^2$ over that set. In other words, $|\psi|^2$ is the probability distribution for the position of that particle. Classically, the time evolution of the classical particle is given by the billiard flow Φ_t described earlier, while the quantum-mechanical time evolution is given by Schrödinger's equation (after setting some physical constants equal to 1),

$$\mathrm{i}h\frac{\partial\psi}{\partial t} = \frac{h^2}{2}\Delta\psi := -\frac{h^2}{2}\left(\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2}\right), \tag{2.1}$$

together with a boundary condition, such as the Dirichlet condition (vanishing at the boundary of ∂B) to ensure self-adjointness. (Self-adjointness is important because it means that the solution operator $U(t)$, which takes the initial condition to the solution at time t, is unitary. This means that the L^2 norm of ψ is constant in time, which is essential for the interpretation of $|\psi|^2$ as a probability density.) Here h, Planck's constant, is for us just a small parameter.

Equation (2.1) has a solution in terms of an infinite sum over the eigenfunctions of Δ. The operator Δ, with the Dirichlet boundary condition, has a sequence of eigenfunctions ϕ_j, with eigenvalues $E_j \to \infty$, that form an orthonormal basis of $L^2(B)$. We can expand any L^2 function ψ in this basis. In particular, we can expand the wavefunction at time $t = 0$:

$$\psi(x, 0) = \sum_j a_j\phi_j(x).$$

Then the time evolution under (2.1) is

$$\psi(x, t) = \sum_j \mathrm{e}^{\mathrm{i}tE_j/h}a_j\phi_j(x). \tag{2.2}$$

---

[a]Strictly, a cotangent vector; however, in $\mathbb{R}^2$, tangent and cotangent vectors can be identified.

Now we can ask: what are the quantum analogues of ergodicity and unique ergodicity? That is, what should be the definitions of quantum ergodicity (QE) and quantum unique ergodicity (QUE)?

Let us first take unique ergodicity. If (X, μ, T) is uniquely ergodic, then every orbit is dense. To see this, suppose, for a contradiction, that the orbit of point x were not dense. Then the closure V of that orbit would be an invariant set, and we could form an invariant measure on V by taking limits of atomic measures supported on $T^{-m}x, T^{-m+1}x, \ldots, T^m(x)$ each with mass $1/(2m + 1)$; these would have a weak limit ν invariant under T. Such a measure ν would be supported away from an nonempty open set, namely the complement of V. It would therefore be different from μ (which is by assumption positive on each nonempty open set), but this contradicts unique ergodicity.

The classical limit of quantum mechanics arises by taking the limit as Planck's constant h tends to zero[b]. Now suppose (for the sake of exposition) that we are interested in our quantum system at a fixed range of energies, say between E and $2E$. The energy of the system is given by the Hamiltonian, in our case $h^2\Delta$. So this means that we are considering states which are linear combinations of eigenfunctions of $h^2\Delta$ with eigenvalues in the range $[E, 2E]$, or equivalently eigenvalues of Δ in the range $[h^{-2}E, 2h^{-2}E]$ — that is, larger and larger eigenvalues, as $h \to 0$. Let us ask: what does it mean that every quantum orbit becomes equidistributed, in the limit $h \to 0$? Suppose a sequence h_j is given, tending to zero, with corresponding initial states ψ_j formed of eigenfunctions with eigenvalues in the range $[h^{-2}E, 2h^{-2}E]$; then it should mean that the long-run average probability distribution,

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} |\psi_j(x, t)|^2 \, \mathrm{d}t$$

converges, as $j \to \infty$, to the uniform distribution over the billiard B. Convergence is meant here in the weak-$*$ sense (see Remark 2.1). If this were true for all sequences of h_j and ψ_j, it would certainly imply that the eigenfunctions themselves become equidistributed, since if the initial state ψ_j is a single eigenfunction, with eigenvalue $h_j^{-2}E_j$, then its time evolution is simply

$$\psi_j(x, t) = \mathrm{e}^{\mathrm{i}E_j t/h_j} \psi_j(x, 0) \implies |\psi_j(x, t)|^2 \text{ is independent of } t.$$

Thus its long-run average probability distribution is just that of the eigenfunction itself. It turns out that the converse is also true: if the eigenfunctions become uniformly distributed[c] as the eigenvalue tends to infinity, then any sequence of eigenfunctions ψ_j made up of eigenfunctions with eigenvalues in the range $[h_j^{-2}E, 2h_j^{-2}E]$, $h_j \to 0$, has a long-run average probability distribution that tends to the uniform distribution in the limit $j \to \infty$.

---

[b]In physics, h has a particular value; however, it has units of $(\mathrm{mass})(\mathrm{length})^2(\mathrm{time})^{-1}$ and therefore may be either small or large in the units given by the typical mass, length and time-scales in a particular situation. One sees classical-type behaviour in situations where h is very small expressed in these units.

[c]If there are multiple eigenvalues then this condition has to be taken in the strong sense that any choice of basis for the eigenspaces leads to uniform distribution.

Remark 2.1. Recall that a sequence of measures ν_j converges in the weak-$*$ sense to ν if for all continuous functions f we have

$$\lim_{j\to\infty} \int f \, \mathrm{d}\nu_j = \int f \, \mathrm{d}\nu.$$

That this is the right notion of convergence becomes clear when considering a simple one-dimensional problem, where the eigenfunctions are $\sqrt{2/\pi} \sin nx$ on $[0,\pi]$, say, the probability distributions are $2/\pi \sin^2 nx$, and these converge weak-$*$ to the uniform distribution on $[0,\pi]$. This is a useful mental picture to keep in mind: typically eigenfunctions with large eigenvalue are highly oscillatory (as in Figures 3 and 4), but the weak-$*$ limit can nevertheless be quite smooth.

Now we can give a rough definition of QUE: a billiard system with L^2-normalised eigenfunction sequence $\phi_1, \phi_2, \ldots$ (ordered by increasing eigenvalue) is QUE if the sequence of probability measures $|\phi_j|^2$ converges in the weak-$*$ sense to the uniform measure on B. The fully-fledged definition of QUE actually is more elaborate, and involves equidistribution not just in space but also in the momentum variables. This requires use of either the Bargmann transform or pseudodifferential operators to obtain measures on phase space rather than just physical space; describing this is beyond the scope of this article. See for example [5].

3. Quantum ergodicity

Now that we have defined QUE, we can define the related notion of quantum ergodicity (QE). As with QUE it is supposed to be a close analogue of the corresponding classical property, ergodicity. In terms of the orbits of points, ergodicity is equivalent to the equidistribution of almost every orbit (see (1.1)). Correspondingly, QE is related the the equidistribution of almost every quantum orbit. As with QUE it has a description in terms of eigenfunctions. It is as follows: we say that a billiard system is QE if there is a subset of positive integers J of density 1, that is, with

$$\lim_{N\to\infty} \frac{\#J \cap \{1,2,\ldots,n\}}{N} = 1,$$

such that the sequence of probability measures $|\phi_{n_j}|^2$, with the increasing sequence n_j restricted to lie in J, converges in the weak-$*$ sense to the uniform measure on B. Let us call the eigenfunctions ϕ_m with $m \notin J$ exceptional. Then we can think heuristically of this condition as follows: when h is sufficiently small, this condition is saying that only a very small proportion of the eigenfunctions with eigenvalues in the range $[h^{-2}E, h^{-2}2E]$ are exceptional. Therefore, a random linear combination of these eigenfunctions will have (with high probability) only a very small component formed from exceptional eigenfunctions, and therefore will be (with high probability) close to being equidistributed. Thus, as $h \to 0$, we can say in a sense that 'almost every' quantum orbit is equidistributed.

4. Theorems

So what is known about QE and QUE?

The first major theorem proved in this subject was the quantum ergodicity theorem proved by Schnirelman, Zelditch and Colin de Verdière for manifolds without boundary (Example 1.3), and for manifolds with boundary by Gérard–Leichtnam and Zelditch–Zworski [11], [15], [4], [6], [16].

Theorem 4.1 (Quantum ergodicity theorem)**.** *If a classical system (in the context of Examples 1.2 and 1.3) is ergodic, then the corresponding quantum system is QE.*

In particular, Examples 1.5 and 1.6 are QE.

So what about QUE? Is there a correspondingly close relationship between classical and quantum unique ergodicity? In fact, heuristically one does not expect such a close relationship. The reason for this is that classical unique ergodicity is an extremely strong condition — the existence of a single periodic orbit is enough to show that classical unique ergodicity fails. On the other hand, quantum mechanics is not expected to be so sensitive to individual orbits. Since quantum particles are somehow 'smeared out', the behaviour of a quantum state at a periodic orbit can be expected to depend crucially on the behaviour of the flow in a neighbourhood of that orbit, that is, on the *stability* of the orbit. If the orbit is stable, then a quantum state can indeed concentrate near a single orbit; this happens for example on the disc (Example 1.4). If it is strongly unstable (as is the case for Anosov systems), then nearby orbits will diverge at an exponential rate, either as time tends to $+\infty$ or $-\infty$, and this is likely to make it difficult for a smeared-out quantum orbit to remain close to the periodic classical orbit. Indeed, for this reason (and because of evidence from their study of arithmetic surfaces) Rudnick and Sarnak made a now-famous conjecture, that the unit tangent bundle of any compact hyperbolic manifold (which are strongly chaotic, that is, Anosov, and all of which are known to be ergodic) is QUE [12]. This is in spite of the fact that there are many invariant classical Borel measures on such spaces: for example, there are known to be many periodic orbits, the number of which grows exponentially as a function of the length.

This conjecture is likely to be very difficult to tackle. However, it has been solved in the case of arithmetic surfaces by Lindenstrauss [10] in his prizewinning work:

Theorem 4.2 (E. Lindenstrauss [10])**.** *QUE holds[d] for arithmetic surfaces, that is, in the case that M is the quotient of hyperbolic space by a congruence subgroup[e].*

This is a consequence of a more general result of Lindenstrauss, in which he classifies measures on certain spaces invariant under certain sorts of group actions; see [10]. The reason that the arithmetic case is more tractable is because, in this case, there are additional symmetries, that is, so-called Hecke operators that commute with the Laplacian. Since these operators commute, one can choose a basis of

---

[d]Technically what is proved here is 'arithmetic QUE', where one studies only joint eigenfunctions of the Laplacian and the Hecke operators; see the discussion below the theorem statement.

[e]An example of a congruence subgroup is $\{M \in SL(2, \mathbb{Z}) \mid M$ is congruent to the identity modulo N, for some fixed $N > 1\}$.

eigenfunctions of the Laplacian that are also eigenfunctions of the Hecke operators (this is significant when there are eigenvalue multiplicities), and one can say more about such joint eigenfunctions.

There has also been progress on nonarithmetic hyperbolic manifolds, and even on manifolds with negative curvature. The culmination of such work is a result by Anantharaman and Nonnenmacher [1], which says that on hyperbolic manifolds, the entropy of any quantum limit is at least half that of the uniform measure, which is a quantitative statement that quantum limits cannot be too localised. However, it does not rule out quantum limits having some highly localised ergodic components.

There have also been many numerical studies of QE and QUE. One that I would like to mention is Barnett's study of Example 1.5, involving accurate computation of about 30 000 eigenfunctions up to about the 700 000th [2]. His results suggest that QUE very likely holds for this billiard, even though, in this case as with hyperbolic manifolds, there are plenty of invariant classical measures that are not equidistributed. Figure 3 shows one numerically computed eigenfunction on this billiard, illustrating that the position probability distribution is random, with no preferred location or direction of oscillation evident.
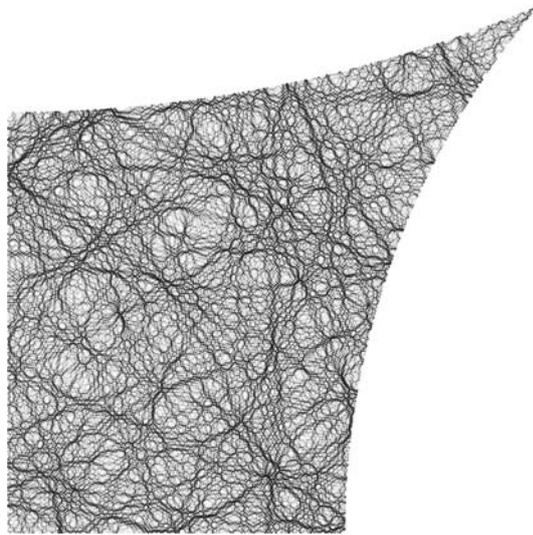

Figure 3. An eigenfunction on the Barnett stadium (Example 1.5).

There have also been many numerical studies of the stadium billard, and there one draws the opposite conclusion: there seem to be sequences of exceptional eigenfunctions that are not equidistributed, but concentrate in the central rectangle. See Figure 4. This was first conjectured by Heller and O'Connor in 1988 [9]. Recently I proved this conjecture, at least for 'almost every' stadium:

Theorem 4.3 (Hassell [7])**.** *Consider the family of stadium domains with central rectangle having dimensions $1 \times t$, for $t > 0$. Then for almost every t, this billiard*
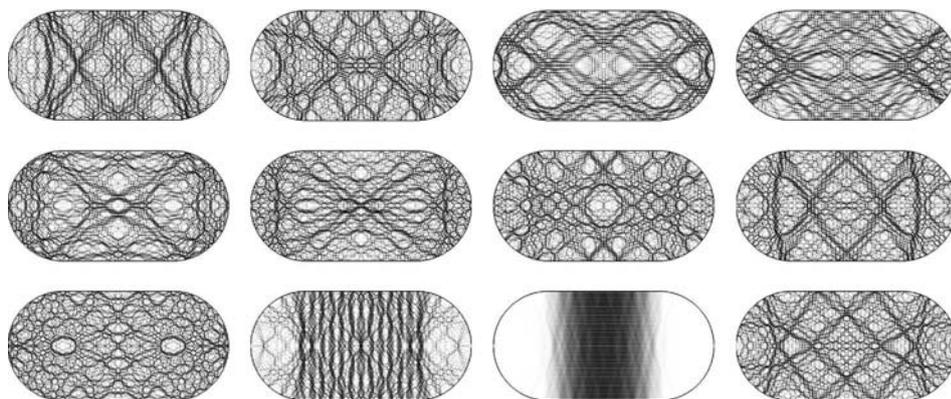
Figure 4. Eigenfunctions on the stadium; the second last is a 'bouncing ball mode' and is not equidistributed.

is not QUE. In fact, for almost every t there are quantum limits that give positive mass to the bouncing ball trajectories (which form a set of measure zero in the uniform measure).

This was the first example of an ergodic billiard proved to be non-QUE. Of course we know that it is QE due to Theorem 4.1, so the sequence of eigenfunctions in Theorem 4.3 is definitely atypical: most eigenfunctions become equidistributed as the eigenvalue tends to infinity. With the help of Luc Hillairet, this result was extended to all partially rectangular billiards (those possessing a central rectangle or cylinder) [7, Appendix]. However, there is plenty more to be understood about eigenfunctions on the stadium. For example, it is not known whether there are sequences of eigenfunctions that concentrate completely in the central rectangle, although the numerical evidence is fairly convincing that such sequences exist. This question is one of my current research interests.

Acknowledgements

References

[1] Anantharaman, N. and Nonnenmacher, S. (2007). Half-delocalization of eigenfunctions for the Laplacian on an Anosov manifold. *Ann. Inst. Fourier (Grenoble)* **57,** 2465–2523.

[2] Barnett, A.H. (2006). Asymptotic rate of quantum ergodicity in chaotic Euclidean billiards. *Commun. Pure Appl. Math.* **59,** 1457–1488.

[3] Bunimovich, L. (1979). On the ergodic properties of nowhere dispersing billiards. *Commun. Math. Phys.* **65,** 295–312.

[4] Colin de Verdière, Y. (1985). Ergodicité et fonctions propres du laplacien. *Commun. Math. Phys.* **102,** 497-502.

[5] Evans, L.C. and Zworski, M. Lectures on Semiclassical Analysis, version 0.8. http://math.berkeley.edu/~zworski/semiclassical.pdf, accessed 18 May 2011.

[6] Gérard, P. and Leichtnam, E. (1993). Ergodic properties of eigenfunctions for the Dirichlet problem. *Duke Math. J.* **71,** 559–607.

[7] Hassell, A. (2010). Ergodic billiards that are not quantum unique ergodic (with an appendix by A. Hassell and L. Hillairet). *Ann. Math.* **171,** 605–618.

[8] Hassell, A. and Zelditch, S. (2004). Quantum ergodicity of boundary values of eigenfunctions. *Commun. Math. Phys.* **248,** 119–168.

[9] Heller, E.J. and O'Connor, P.W. (1988). Quantum localization for a strongly classically chaotic system. *Phys. Rev. Lett.* **61,** 2288–2291.

[10] Lindenstrauss, E. (2006). Invariant measures and arithmetic quantum ergodicity. *Ann. Math.* **163,** 165–219.

[11] Schnirelman, A. (1974). Ergodic properties of eigenfunctions. *Uspekhi Math. Nauk.* **29,** 181–182.

[12] Rudnick, Z. and Sarnak, P. (1994). The behaviour of eigenstates of arithmetic hyperbolic manifolds. *Commun. Math. Phys.* **161,** 195–213.

[13] Sinai, Ja. G. (1970). Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards. *Uspehi Math. Nauk* **25,** 141–192 (in Russian); *Russ. Math. Surv.* **25,** 137–189.

[14] Stein, E.M. and Shakarchi, R. (2003). *Fourier Analysis: An Introduction.* Princeton University Press, Princeton, NJ.

[15] Zelditch, S. (1987). Uniform distribution of eigenfunctions on compact hyperbolic surfaces. *Duke Math. J.* **55,** 919–941.

[16] Zelditch, S. and Zworski, M. (1996). Ergodicity of eigenfunctions for ergodic billiards. *Commun. Math. Phys.* **175,** 673–682.

Andrew Hassell was an undergraduate at the University of Western Australia and at the Australian National University. He obtained his PhD at the Massachusetts Institute of Technology under the supervision of Richard Melrose. After a two-year postdoc at Stanford University, he returned to the ANU in 1996 on an ARC postdoctoral fellowship and has been there ever since. Currently he is an Associate Professor and ARC Future Fellow. Andrew's research interests are in partial differential equations, spectral theory, harmonic analysis and quantum ergodicity. He was awarded the Australian Mathematical Society Medal in 2003.