PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 102, Number 1, January 1988

ON THE SPHERE CONJECTURE OF BIRKHOFF

RICHARD JERRARD

(Communicated by George R. Sell)

ABSTRACT. Birkhoff's sphere conjecture, now known to be false, says that if f is a measure preserving homeomorphism of S^2 with the poles N and S fixed, and with no other periodic points, then f is topologically conjugate to an irrational rotation of S^2. In this setting we say that $D \subset S^2$ is *maximal for* f if $f(D) \cap D = \emptyset$ and D is maximal with respect to that property. Also, f is *2-small* if for any circular ball B such that $f(B) \cap B = \emptyset$, $f^{-1}(B) \cap f(B) = \emptyset$ also.

THEOREM. *Any f as above and also 2-small has a maximal set D which is an open ball; its boundary contains N and S and is locally connected, and the area of D is an irrational fraction of the area of S^2. This theorem gives another way of looking at the maps involved in the Birkhoff conjecture.*

1. Introduction. We want to consider G. D. Birkhoff's sphere conjecture. It was one of sixteen questions raised in a paper presented by him in Chicago in September, 1941, which appeared only in summary form in Science [**1**].

BIRKHOFF'S CONJECTURE. If f is an orientation and measure preserving homeomorphism of the two-sphere S^2 with exactly two fixed points but no other periodic points, then f is topologically conjugate to an irrational rotation of S^2. We will call any such map a *Birkhoff map.*

A partial analysis of this problem was given by Montgomery [**5**]. However, the conjecture is now known to be false. One counterexample can be obtained by varying an example of a sphere mapping given by Handel [**2**] in 1982. More recently Markus [**4**] has shown that the conjecture is true for such a homeomorphism that is additionally the unit-time map for a smooth, conservative flow on S^2.

In this paper we obtain a result indicating in a different way how the idea of the conjecture can be interpreted. We use the following definition.

DEFINITION 1.1. Suppose that f is a homeomorphism on a topological space X. A *maximal set for* f is any connected, open set $D \subset X$ such that $f(D) \cap D = \emptyset$, and which is maximal with respect to that property. It then follows that $f^{-1}(D) \cap D = \emptyset$ also.

For example, if h is a rotation of the sphere, one maximal set is a lune from one fixed point to the other. We will show here that for many Birkhoff maps there exists a maximal set that is similar to such a lune. Furthermore, maximal sets are interesting in their own right. If we consider the iterates $f^n(D)$ of a maximal set D we find that either they cover S^2 or $\mathrm{Bd}[\bigcup_n f^n(D)]$ contains a minimal set. The

Received by the editors August 5, 1986.
1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 54H20.
Key words and phrases. Sphere homeomorphism, area preserving.

0002-9939/88 $1.00 + $.25 per page

nature of D seems to depend crucially on whether its boundary is locally connected; both kinds of maximal sets may exist for the same function.

We want to make one further restriction on the maps studied here.

DEFINITION 1.2. A homeomorphism f on S^2 is 2-*small* if for any circular ball B such that $B \cap f(B) = \emptyset$, $f^{-1}(B) \cap f(B) = \emptyset$ also. Any rotation of the sphere of angle less than $2\pi/3$ is 2-small. With this definition we can state the theorem about Birkhoff maps.

THEOREM 1.3. *If f is a 2-small orientation and measure preserving homeomorphism of S^2 which fixes the poles N and S and has no other periodic points, then f has a maximal set D which is an open ball.* $\mathrm{Bd}(D)$ *contains N and S and is locally connected, and the area of D is an irrational fraction of the area of S^2.*

2. Maximal sets. We first prove a few facts about maximal sets.

PROPOSITION 2.1. *If f is a homeomorphism on a Hausdorff space X, and is not the identity, then a maximal set for f exists.*

PROOF. We use the set $\mathbf{U}$ of all connected open sets such that if $U \in \mathbf{U}$ then $f(U) \cap U = \emptyset$. $\mathbf{U}$ is nonempty and is ordered by inclusion, and every chain in $\mathbf{U}$ has an upper bound, namely its union. Then by Zorn's Lemma $\mathbf{U}$ contains a maximal element D, and D is a maximal set for f. □

We will usually denote a maximal set by D_0, and put $D_n = f^n(D_0)$.

PROPOSITION 2.2. *Suppose that D_0 is a connected open set in a locally connected space X such that $D_0 \cap f(D_0) = \emptyset$. Then D_0 is a maximal set if and only if*

$$\mathrm{Bd}(D_0) \subset \mathrm{Bd}(D_{-1}) \cup \mathrm{Bd}(D_1).$$

PROOF. Assume that D_0 is a connected open set with $D_0 \cap f(D_0) = \emptyset$, and that $x \in \mathrm{Bd}(D_0)$. First, if $x \notin \mathrm{Bd}(D_{-1}) \cup \mathrm{Bd}(D_1)$, there is a neighborhood U of x meeting neither D_{-1} nor D_1 such that $f(U) \cap U = \emptyset$ [if $f(x) = x$ then $x \in \mathrm{Bd}(D_n)$ for all n]. We see that

$$(D_0 \cup U) \cap f(D_0 \cup U) = (D_0 \cup U) \cap [f(D_0) \cup f(U)] = \emptyset$$

because $D_0 \cap f(D_0) = U \cap f(U) = \emptyset$, $U \cap f(D_0) = U \cap D_1 = \emptyset$, and $D_{-1} \cap U = \emptyset$ implies $D_0 \cap f(U) = \emptyset$. This property of $D_0 \cup U$ shows that D_0 is not maximal, a contradiction.

Second, suppose that $\mathrm{Bd}(D_0) \subset \mathrm{Bd}(D_{-1}) \cup \mathrm{Bd}(D_1)$ but that D_0 is not maximal. Then D_0 is properly contained in some connected open D_0' and $D_0' \cap f(D_0') = \emptyset$. There is then some $x \in \mathrm{Bd}(D_0) \cap D_0'$. Now $x \in D_0'$ implies $x \notin \overline{D}_1'$ and $x \in \mathrm{Bd}(D_0)$ implies $x \in \overline{D}_1 \subset \overline{D}_1'$, a contradiction. □

We next look at the behavior of the boundary points of a maximal set D_0.

DEFINITION 2.3. We put

$$\begin{aligned} B_0 &= \{x \in \mathrm{Bd}(D_0) | x \notin \mathrm{Bd}(D_1)\}, \\ B_1 &= \{x \in \mathrm{Bd}(D_0) | x \notin \mathrm{Bd}(D_{-1})\}, \\ B_p &= \{x \in \mathrm{Bd}(D_0) | x \in \mathrm{Bd}(D_{-1}) \cap \mathrm{Bd}(D_1)\} \\ &= \mathrm{Bd}(D_{-1}) \cap \mathrm{Bd}(D_0) \cap \mathrm{Bd}(D_1). \end{aligned}$$

In the example where h is a small rotation of S^2 with N and S fixed, if the maximal set is a lune from N to S then $B_p = \{N, S\}$, B_1 and B_0 are the leading and trailing edges of the lune, $h(B_0) = B_1$, and $h(B_p) = B_p$. We shall see that this is a typical case. However, if h is a rotation by π radians, then D_0 is a hemisphere and B_p is its entire great circle boundary. We will require h to be 2-small to rule out this situation.

EXAMPLE 2.4. Following Kerékjártó [3] we obtain a more disquieting example of a maximal set for the same 2-small rotation h of S^2. Suppose that S^2 is described by the equation $x^2+y^2+z^2 = 1$ and that α: $(0, 1) \to S^2$ is a path that is asymptotic as $t \to 0$ $(t \to 1)$ to the circle $z = -1/2$ $(z = 1/2)$, and which is strictly monotone increasing in its z coordinate. Take β to be the path obtained by following α by h. The two paths α and β each encircle S^2 infinitely many times. Choose for D_0 the narrow open set enclosed by the images of α and β; then B_0 is the image of α, B_1 is the image of β and B_p is the union of the two circles $z = \pm 1/2$. Again $h(B_p) = B_p$.

A maximal set of this sort gives no information about the map h for $z > 1/2$ or $z < -1/2$. Note that along the circular boundaries $\mathrm{Bd}(D_0)$ is not locally connected. The two kinds of maximal sets are fundamentally different.

Now suppose that D_0 is a maximal set for f on S^2 and consider the sets $F = \mathrm{Cl}[\bigcup\{D_n | n \in Z\}]$ and $G = S^2 - F$. We show that G is f-invariant.

PROPOSITION 2.5. $f(G) = G$.

PROOF. Suppose there exists $x \in f(G)$ such that $x \notin G$. Since G is open there is a ball $B(x, \delta) \subset f(G)$ such that $B \cap G = \emptyset$. Now $B \cap D_m \neq \emptyset$ for some m, so $f^{-1}(B) \cap D_{m-1} \neq \emptyset$. But $f^{-1}(B) \subset G$ and $G \cap D_{m-1} = \emptyset$, and this contradiction shows that $f(G) \subset G$. In the same way it follows that $G \subset f(G)$. □

PROPOSITION 2.6. $\mathrm{Bd}(G)$ *contains a minimal set.*

PROOF. We need only note that $f(\mathrm{Bd}(G)) = \mathrm{Bd}(G)$, and that $\mathrm{Bd}(G)$ is closed. Then any $x \in \mathrm{Bd}(G)$ generates a minimal set. □

In Example 2.4 $G = \{(x, y, z) \in S^2 | z < -1/2 \text{ or } z > 1/2\}$ and $\mathrm{Bd}(G)$ consists of the two circles $z = \pm 1/2$. Both circles are minimal sets if the rotation is irrational and contain finite minimal sets if it is rational. We want to find maximal sets where this situation does not occur. In the following theorem we construct maximal sets with locally connected boundaries.

THEOREM 2.7. *If f is a 2-small homeomorphism on S^2 other than the identity, then there is a maximal set D_0 such that* $\mathrm{Bd}(D_0)$ *is locally connected, except possibly at fixed points of f.*

PROOF. (a) We show that $D_{-1} \cap D_1 = \emptyset$.

We will construct our maximal set D_0 for f as a union of open sets

$$D_0 = \bigcup_n [B_{0n} \cup R_{0n}].$$

Each B_{0n} is a circular disc such that $B_{0n} \cap f(B_{0m}) = \emptyset$ for all $m \leq n$; because f is 2-small we have $f(B_{0n}) \cap f^{-1}(B_{0n}) = \emptyset$. The sets R_{0n} are chosen so that R_{-1n}, R_{0n}, and R_{1n} are disjoint. Therefore the maximal set D_0 will have the crucial property that $D_{-1} = \bigcup_n f^{-1}(B_{0n} \cup R_{0n})$ and $D_1 = \bigcup_n f(B_{0n} \cup R_{0n})$ are disjoint.

(b) $\mathrm{Bd}(D_0) \cap \mathrm{Bd}(D_1)$ is connected.

Suppose that $\mathrm{Bd}(D_0) \cap \mathrm{Bd}(D_1)$ is not connected. Then the connected set $\overline{D}_0 \cup \overline{D}_1$ separates S^2. Since $\mathrm{Bd}(D_0) \subset \mathrm{Bd}(D_{-1}) \cup \mathrm{Bd}(D_1)$, each component of $S^2 - (\overline{D}_0 \cup \overline{D}_1)$ contains points of D_{-1}. But $D_{-1} \cap (\overline{D}_0 \cup \overline{D}_1) = \emptyset$, so D_{-1} is not connected, a contradiction.

(c) If $\mathrm{Bd}(D_0)$ is not locally connected at x, then $x \in B_p$.

First suppose that a maximal set D_0 such that $D_{-1} \cap D_1 = \emptyset$ exists. We show how to alter D_0 so that if $\mathrm{Bd}(D_0)$ is not locally connected at x, then x is a common boundary point of D_{-1}, D_0 and D_1, that is, $x \in B_p$. To see this, suppose that $x \in \mathrm{Bd}(D_0) \cap \mathrm{Bd}(D_1)$ but $x \notin \mathrm{Bd}(D_{-1})$, and that $\mathrm{Bd}(D_0)$ is not locally connected at x. Choose a small open disc $B(x, \rho)$ not intersecting $\overline{D}_{-1}$. We first add to D_1 the set $A = B(x, \rho) \cap D_0$. This requires that $f(A)$ be added to D_2 (and hence subtracted from D_1) etc. We obtain $D_1' = [D_1 \cup A] - f(A)$ and $D_0' = [D_0 \cup f^{-1}(A)] - A$. Due to the subtraction of A, D_0' is no longer connected, but only one component of $\overline{D}_0'$ meets $\overline{D}_{-1}$ for $\mathrm{Bd}(D_0) \cap \mathrm{Bd}(D_{-1})$ is connected. The other components of D_0' are completely surrounded by D_1; we add all of their closures to D_1' to get the new D_1, with the corresponding new D_{-1} and D_0. We have altered D_0 and D_1 so that the set within $B(x, \rho)$ where their common boundary was not locally connected is eliminated, and no new points where the boundary is not locally connected are created. A finite number of such operations (the ball $B(x, \rho)$ may be replaced by a topological ball) will produce the desired D_0. We will henceforth assume that if $\mathrm{Bd}(D_0)$ is not locally connected at x, then $x \in B_p$.

(d) *Construction of* D_0. We will obtain D_0 as a union of open sets which are constructed inductively. To begin, choose a point $P \in S^2$ such that $f(P) \neq P$, and denote by B_{00} the maximal open disc with center at P such that $B_{00} \cap B_{10} = \emptyset$ (in this notation the first subscript indicates how many times f has been applied: $f(B_{00}) = B_{10}$; the second labels the inductive step). Note that because f is a homeomorphism, the intersection $\overline{B}_{00} \cap \overline{B}_{10}$ corresponds exactly to $\overline{B}_{-10} \cap \overline{B}_{00}$. Now, if there is an open set R_{00} bounded by $\overline{B}_{00} \cup \overline{B}_{10}$, and not containing B_{-10} we unite B_{00} with R_{00} and define $D_{01} = \mathrm{Int}[\overline{B_{00} \cup R_{00}}]$, the first in the sequence of open sets $\{D_{0n}\}$. The *free boundary* of D_{01}, namely $\mathrm{Bd}(D_{01}) - [\mathrm{Bd}(D_{-11}) \cup \mathrm{Bd}(D_{11})]$, consists of two circular arcs a_{01} and b_{01}. The open sets D_{-11}, D_{01} and D_{11} are pairwise disjoint, and $\mathrm{Bd}(D_{01})$ consists of the two circular arcs plus points in $\mathrm{Bd}(D_{-11}) \cup \mathrm{Bd}(D_{11})$.

The second step is typical (see Figure 1). We choose another point P on the free boundary of D_{01} equidistant from D_{-11} and D_{11} and such that this distance is maximal. The maximal open disc B_{01}, centered at P and such that $(D_{01} \cup B_{01}) \cap (D_{11} \cup B_{11}) = \emptyset$, is added to D_{01} to get $D_{02}' = D_{01} \cup B_{01}$. Note that $\overline{B}_{01}$ intersects both $\overline{D}_{-12}'$ and $\overline{D}_{12}'$, and if it intersects $\overline{B}_{11}$ then it must also intersect $\overline{B}_{-11}$. Now, if there is an open set R_{01} bounded by $\overline{D}_{02}' \cup \overline{D}_{12}'$ (and not containing D_{-12}') we join it to D_{02}' to get the second set in the sequence, $D_{02} = \mathrm{Int}[\overline{D_{02}' \cup R_{01}}]$. The free boundary of D_{02} again consists of two circular arcs a_{02} and b_{02}.

We can imagine that the disc B_{01} grows from its center at P until it strikes either the growing image disc B_{11} or the fixed set D_{11}. The first possibility is illustrated in Figure 1. In the other case the growing B_{01} must hit D_{11} and D_{-11} simultaneously (because P is equidistant from these sets) and then the growing

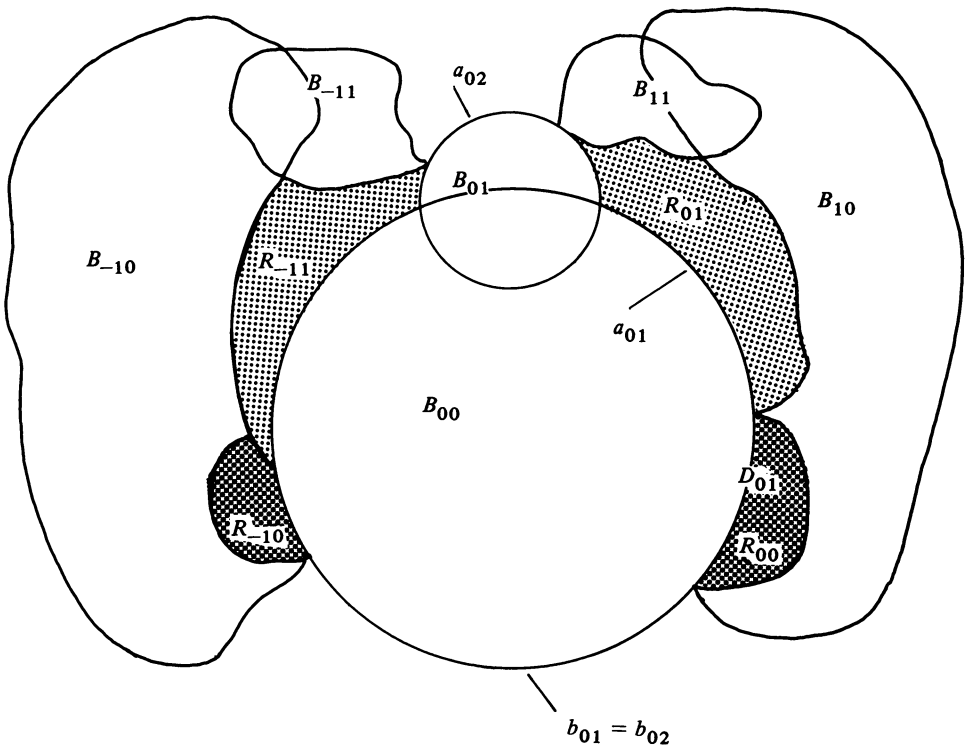

FIGURE 1. First stages in the construction of D_0

images B_{-11} and B_{11} will simultaneously hit B_{00}. If P is on a_{01} then $\overline{B}_{-11}$ and $\overline{B}_{11}$ will meet $\overline{B}_{00}$ on b_{01}. The result in both cases is that a_{01} is buried and a new circular free boundary arc a_{02} for D_{02} is created, but that either all of b_{01} or a subarc of it survives to become b_{02}.

The construction continues in this way to yield a sequence $\{D_{0n}\}$ of open sets, each one contained in its successor. We define $D_0 = \bigcup_n D_{0n}$. This is clearly a connected open set such that $D_0 \cap f(D_0) = \varnothing$. It is maximal because $\mathrm{Bd}(D_0) \subset [\mathrm{Bd}(D_{-1}) \cup \mathrm{Bd}(D_1)]$. To see this first note that the sequence of radii of added discs converges to zero, for no added disc can have its center in a preceding disc, and the area of S^2 is finite. Then we can add no disc on $\mathrm{Bd}(D_0)$, so each point of $\mathrm{Bd}(D_0)$ must also be in $\mathrm{Bd}(D_{-1}) \cup \mathrm{Bd}(D_1)$. It follows from Proposition 2.2 that D_0 is a maximal set. This argument is unaffected if we add open discs B and associated newly enclosed regions R in some other order, which it will be necessary to do.

(e) D_{0m} cannot separate S^2.

At this stage we can see that if for some m, D_{0m} separates S^2, then $\mathrm{Bd}(D_0)$ is locally connected. For if D_{0m} separates S^2 so does D_0, and then $\mathrm{Bd}(D_0)$ has more than one component. Since D_{-1} and D_1 are both connected, and since they both have boundary points in common with D_0, they must lie in different components of $(S^2 - D_0)$. Then D_{-1}, D_0 and D_1 have no common boundary point, and $\mathrm{Bd}(D_0)$ must be locally connected.

(f) *Proof that* $\mathrm{Bd}(D_0)$ *is locally connected.* In outline, the argument here goes as follows. We have seen that at each stage of the construction there are just two free circular arcs on the boundary on which discs may be added. We change the order of construction and add a complete sequence of discs $\{B_{0n}\}$ continually on one of the arcs a_{0n}, leaving b_{0n} possibly shortened but otherwise undisturbed. If this sequence $\{B_{0n}\}$ converges to a point x at which the boundary is not locally connected, and which is not a fixed point, then each member of the sequence $\{B_{1n}\}$ must touch b_{0n}; it must touch one of the two free arcs, and because x is not a fixed point, B_{1n} is relatively far from a_{0n}. It then follows that $f(x) \in b_{0n}$, and as soon as we start adding discs on b_{0n} we arrive at a contradiction.

We first alter the construction by adding discs and regions in a different order. We first construct a sequence of discs B_{0n} with centers always on the free boundary arcs a_{0n}, together with their associated regions R_{0n}, to get the sequence $\{D_{0n}\}$. We put $D_0' = \bigcup_n D_{0n}$. During this process we have $b_{01} \supset b_{02} \supset b_{03} \supset \cdots$. No disc is added with center on the second free boundary arc, though it may be shortened when R_{0n} is added; thus b_{0n} is a subarc of b_{01}. If during this process some added disc B_{0m} intersects b_{0m} then D_{0m+1} separates S^2 and we can apply the previous paragraph. D_0' has only one free boundary arc, as one has been eliminated by adding this first sequence of discs and regions. The single circular free boundary segment remaining in D_0' we will call c_{00}.

We next add successively to D_0' maximal discs C_{0n} centered on the free boundary arcs c_{0n} together with the associated newly enclosed regions S_{0n} to get an increasing sequence of open sets E_{0n}. Here C_{0n} and S_{0n} correspond to B_{0n} and R_{0n} in the earlier sequence. We put $E_{01} = D_0' \cup \mathrm{Int}[\overline{C_{00} \cup S_{00}}]$, and $E_{0n+1} = E_{0n} \cup \mathrm{Int}[\overline{C_{0n} \cup S_{0n}}]$. The center of C_{0n} is at a point $P \in c_{0n}$. Finally define $D_0 = D_0' \cup [\bigcup_n E_{0n}]$; as before, D_0 is a maximal set.

Now suppose that at some y such that $f(y) \neq y$, $\mathrm{Bd}(D_0)$ is not locally connected. We will show first that either y or $x = f^{-1}(y)$ is a limit point of a subsequence of one of the two sequences of added discs, and second that this fact leads to a contradiction.

For the first part, if y is not a limit point of a sequence of added discs then it must be a limit point of a subsequence of a sequence of added regions, say of the sequence $\{R_{0n}\}$. That is, each ε-ball centered at y intersects infinitely many of the regions $\{R_{0n}\}$. In particular it must intersect infinitely many of the boundaries $\{\mathrm{Bd}(R_{0n})\}$. But $\mathrm{Bd}(R_{0n})$ is a finite union of segments of disc boundaries, $\mathrm{Bd}(B_{0n})$ and $\mathrm{Bd}(B_{0n-1})$, and of homeomorphic images of disc boundaries, $f[\mathrm{Bd}(B_{0n-1})]$ and $f[\mathrm{Bd}(B_{0n})]$ (see R_{01} in Figure 1). Therefore, either every neighborhood of y intersects (and in fact contains) infinitely many discs B_{0n} or every neighborhood of $x = f^{-1}(y)$ does so. Thus we can conclude that either x or y is a limit point of a subsequence of one of the sequences of added discs.

For the second part, using the preceding paragraph we will assume that at some x with $f(x) \neq x$, $\mathrm{Bd}(D_0)$ is not locally connected and every neighborhood of x contains infinitely many of the added discs $\{B_{0n}\}$ (we shall see presently that x cannot be a limit point of the second sequence $\{C_{0n}\}$). Recall that each disc B_{0n} is added to the free boundary segment a_{0n} and leaves the boundary segment b_{0n} undisturbed. Assume that the limit point x is not on the free boundary segment $c_{00} \subset b_{0n}$ and choose N so that for $n > N$, $\mathrm{diam}(B_{0n})$ is much smaller than the

distances $d[x, f^{-1}(x)]$ and $d[x, f(x)]$. Then for $n > N$, when B_{0n} is added we can imagine that it grows from its center $P \in a_{0n}$ until it simultaneously strikes D_{-1n} and D_{1n}. Far away, B_{-1n} and B_{1n} grow from their centers at $f^{-1}(P)$ and $f(P)$ and must both strike D_{0n} on its free boundary, and the only possibility is that they both strike b_{0n}. We now have a sequence of discs converging to x whose radii converge to zero. The image and counter-image sets all touch b_{0n} and since f is uniformly continuous the diameters of the image and counter-image sets also converge to zero. Therefore $f^{-1}(x)$ and $f(x)$ both lie on b_{01}, one of the original circular boundary segments of D_{01}. Then because $b_{01} \subset \mathrm{Bd}(D_{01})$ we find that $x \in \mathrm{Bd}(D_{11})$ and $x \in \mathrm{Bd}(D_{-11})$, or $x \in \overline{D}_{-11} \cap \overline{D}_{11}$.

Now, in continuing the construction of D_0 we add the first disc C_{00} with center on c_{00}, together with the associated newly enclosed region S_{00}, to D_0' to obtain E_{01}. But each point of the free segment c_{00} is now contained either in the interior of E_{01} or in a locally connected circular boundary segment of E_{01}. The situation is entirely like that in Figure 1, where each point of a_{01} lies either in the interior of D_{02} or on the circular segment that is the common boundary of D_{02} and R_{-11}. Therefore either $f(x) \in \mathrm{Int}(D_0)$, an impossibility because $x \in \mathrm{Bd}(D_{-1})$, or $f(x)$ is on the common boundary of only two of the three sets D_{-1}, D_0 and D_1. This is again impossible because we have shown that $x \in B_p$.

This completes the argument which shows that if D_0 is constructed as above its boundary must be locally connected. It remains to clear up a couple of loose ends. First, the hypothetical point x at which $\mathrm{Bd}(D_0)$ is not locally connected and which is the limit of a sequence of added discs cannot be the limit of the second sequence of added discs $\{C_{0n}\}$. For during this stage in the construction there is only one free boundary segment c_{0n}, and when C_{0n} is added on c_{0n} the corresponding discs $\overline{C}_{-1n}$ and $\overline{C}_{1n}$ must intersect either c_{0n} or $\overline{C}_{0n}$. Therefore, recalling that P is the center of the added disc, as n becomes large $\mathrm{diam}(C_{0n}) \to 0$ and hence $d[P, f(P)] \to 0$ and $d[P, f^{-1}(P)] \to 0$. It follows that x is then a fixed point of f and is excluded from the argument by the hypothesis of the theorem. Second, we assumed after obtaining x that it is not on c_{00}. If it is then we find as above that x, $f^{-1}(x)$ and $f(x)$ are all on $c_{00} \subset b_{01}$. But $x \in \mathrm{Bd}(D_{01})$ implies $f(x) \in \mathrm{Bd}(D_{11})$, and no point of b_{01}, by definition, lies on $\mathrm{Bd}(D_{11})$. □

3. Proof of Theorem 1.3. Using the results of §2 we can prove Theorem 1.3.

LEMMA 3.1. *If f is an area-preserving homeomorphism on S^2 (not the identity) then any maximal set D_0 for f with locally connected boundary is an open disc.*

PROOF. Suppose that D_0 is not a disc, and therefore D_0 separates S^2. Its complement must contain exactly two components, for by Proposition 2.2 its boundary cannot contain more than two components; one of them lies in $\mathrm{Bd}(D_{-1})$ and the other in $\mathrm{Bd}(D_1)$. Then D_1 lies in the component of the complement of D_0 which does not contain D_{-1} and it also separates S^2. Similarly, D_n lies in the component of the complement of D_{n-1} that does not contain D_{n-2}. Thus the sets $D_0, D_1, D_2, \ldots$ are all disjoint from one another, and this is impossible for they all have the same area. □

We recall Definition 2.3 for the following lemma.

LEMMA 3.2. *If f is a 2-small area-preserving homeomorphism of S^2 (not the identity) and D_0 is a maximal set with locally connected boundary, then $f(B_p) = B_p$ and $f(B_0) = B_1$.*

PROOF. We first prove that $f(B_p) = B_p$. We assume that $x \in B_p$ and $f(x) \notin \mathrm{Bd}(D_{-1})$; thus $f(x) \in B_1$. Choose a small neighborhood U of $f(x)$ so that $U \cap \mathrm{Bd}(D_0) \subset B_1$. Then $U \cap \mathrm{Bd}(D_0)$ lies on the common boundary of the two open discs D_0 and D_1, and is locally connected because x is not a fixed point. But $x \in \mathrm{Bd}(D_1)$ implies $f(x) \in \mathrm{Bd}(D_2)$ also, and D_0, D_1 and D_2 are pairwise disjoint because f is 2-small. It follows that either $U \cap D_0$ or $U \cap D_1$ is not connected, say $U \cap D_0$. Then $D_0 \cup \{f(x)\}$ separates S^2, with D_{-1} and D_1 in different components. But the argument of Lemma 3.1 shows that this is not possible.

Now $f(B_p) \subset B_p$ and the same argument shows that $f^{-1}(B_p) \subset B_p$, or $B_p \subset f(B_p)$; hence $B_p = f(B_p)$. It then follows that $f(B_0) = B_1$. □

The following lemma completes the description of the maximal set for the 2-small Birkhoff map, except for the proof that its area is irrational. It consists of the open set between the two arcs B_0 and B_1, which run from N to S with $f(B_0) = B_1$.

LEMMA 3.3. *If f is a 2-small area-preserving homeomorphism of S^2 with fixed points at N and S as the only two periodic points, then any maximal set D_0 with locally connected boundary satisfies $B_p = \{N, S\}$.*

PROOF. By Lemma 3.2, B_p is an invariant set. Since D_0 is an open disc, its locally connected boundary is a topological circle. We know that $\mathrm{Bd}(D_0)$ is the disjoint union of B_0, B_1 and B_p. Further B_0 (and hence B_1) is connected, for otherwise $D_0 \cup D_{-1}$ would separate D_1. Thus B_0 and B_1 are open intervals in $\mathrm{Bd}(D_0)$ and are separated by B_p, which therefore has exactly two components. The two components of B_p contain respectively the fixed points N and S. Finally, each component of B_p must consist of a single point: recalling that B_p is the common boundary of the three open discs D_{-1}, D_0 and D_1, any component of B_p which contains more than one point cannot be locally connected. But we have assumed that $\mathrm{Bd}(D_0)$ is locally connected (and 2.5 assures us that such maximal sets exist). Hence $B_p = \{N, S\}$. □

PROOF OF THEOREM 1.3. By the preceding lemmas we know that there exists a maximal set D_0 for f which is an open disc with the fixed points N and S on its locally connected boundary.

It remains to show that if A_S is the area of S^2 and A_0 is the area of D_0, then A_0/A_S is irrational. We assume that it is rational, that $A_0/A_S = p/q$, and consider the function f^q. This is also an area preserving homeomorphism of S^2 with the fixed points N and S as the only periodic points. It is also easy to see that if f is 2-small then so is f^q. We then know that if $f^q \neq 1$ there exists a maximal set D for f^q which is an open disc with N and S on its locally connected boundary, which in turn consists of two nonintersecting arcs from N to S, one of which is the image under f of the other. Furthermore the area of D must be q times the area of A_0, reduced modulo the area of S^2, that is, $qA_0 \pmod{A_S}$. But $qA_0 \pmod{A_S} = q(pA_S/q) \pmod{A_S} = 0$. Therefore D does not exist and f^q is the identity. Since we assumed only two periodic points for f we conclude that A_0/A_S is irrational. □

REFERENCES

1. G. D. Birkhoff, *Some unsolved problems of theoretical dynamics*, Science **94** (1941), 598–600.
2. Michael Handel, *A pathological C^∞ diffeomorphism of the plane*, Proc. Amer. Math. Soc. **86** (1982), 163–168.
3. B. V. Kerékjártó, *Vorlesungen uber Topologie*. I, Grundlehren Math. Wiss., Bd. VII, Springer-Verlag, Berlin and New York, 1923, p. 195.
4. L. Markus, *Three unresolved problems of dynamics*, Preprint, University of Warwick, 1985.
5. D. Montgomery, *Measure preserving homeomorphisms at fixed points*, Bull. Amer. Math. Soc. **51** (1945), 949–953.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS, URBANA, ILLINOIS 61801