Forming the Borromean Rings out of polygonal unknots

Hugh Howards

1 Introduction

The Borromean Rings in Figure 1 appear to be made out of circles, but a result of Freedman and Skora shows that this is an optical illusion (see [F] or [H]). The Borromean Rings are a special type of Brunninan Link: a link of \( n \) components is one which is not an unlink, but for which every sublink of \( n - 1 \) components is an unlink. There are an infinite number of distinct Brunnian links of \( n \) components for \( n \geq 3 \), but the Borromean Rings are the most famous example.

This fact that the Borromean Rings cannot be formed from circles often comes as a surprise, but then we come to the contrasting result that although it cannot be built out of circles, the Borromean Rings can be built out of convex curves, for example, one can form it from two circles and an ellipse. Although it is only one out of an infinite number of Brunnian links of three

![Figure 1: The Borromean Rings.](image)
components, it is the only one which can be built out of convex components \cite{H}. The convexity result is, in fact a bit stronger and shows that no 4 component Brunnian link can be made out of convex components and Davis generalizes this result to 5 components in \cite{D}.

While this shows it is in some sense hard to form most Brunnian links out of certain shapes, the Borromean Rings leave some flexibility. This leads to the question of what shapes can be used to form the Borromean Rings and the following surprising conjecture of Matthew Cook at the California Institute of Technology.

**Conjecture 1.1.** (Cook) Given any three unknotted simple closed curves in \(\mathbb{R}^3\), they can always be arranged to form the Borromean Rings unless they are all circles.

In this paper we show that any three polygonal unknots, sometimes called stick knots, (consisting of straight edges meeting at a set of vertices) can be used to form the Borromean Rings through rigid transformations of the components in \(\mathbb{R}^3\) together with scaling of \(\mathbb{R}^3\) applied to the individual components. Note that since any knot can be approximated with a polygonal knot that is arbitrarily close to it, any set of three unknots comes arbitrarily close to forming the Borromean rings: even three circles which themselves cannot form the Borromean Rings.

One of the neat things about this theorem and proof is that the theorem is counterintuitive and yet it can be proven using techniques more or less entirely from freshman calculus and first semester undergraduate linear algebra together with simple combinatorial arguments.

## 2 Any three polygonal knots can be used to form the Borromean rings.

Virtually the same construction will work to prove both of the following theorems. The main difference is that the scaling necessary in the first theorem can be omitted in the second one.

**Theorem 2.1.** Let \(K_1\), \(K_2\), and \(K_3\) be three polygonal unknots, then we may form the Borromean rings out of them through rigid motions of \(\mathbb{R}^3\) applied to the individual components together with scaling of one of the components.
Theorem 2.2. Let $K_1$, $K_2$, and $K_3$ be three polygonal unknots, at least one of which is planar, then we may form the Borromean rings out of them through rigid motions of $\mathbb{R}^3$ applied to the individual components.

We will say that a vertex $v$ of a polygonal knot is an extremal vertex if there is some plane which intersects the knot only in $v$ and therefore there is a direction vector with respect to which $v$ is the knot’s unique global maximum.

In topology two manifolds are said to fail to be in general position (or to not intersect transversally) if moving one of the them an arbitrarily small amount changes the intersection topologically, i.e. up to homeomorphism such as a change in the number of components of intersection or topological type, otherwise they are in general position (and they intersect transversally). For example, the $x$ and $y$ axis in $\mathbb{R}^2$ are in general position because they intersect in one point and moving one of them minimally will not change this, but the $x$ and $y$ axis viewed as a subset of $\mathbb{R}^3$ are not in general position because moving one of the axis less than $\epsilon$ can result in the lines becoming disjoint.

To preview the argument and build intuition, we now give an outline of how the knots will be positioned in the proof. We first pick an extremal vertex for each of the three knots and position the knots so that all three extreme vertices are at the origin, but for $K_1$ the pair of edges leaving the extremal vertex are horizontal (lie in the $xy$-plane) and the other two knots have edges leaving the critical vertex that are vertical, lying in the $yz$-plane as in Figure 2. For the vertical knots we position them so that one has its global maximum at the origin and the other has its global minimum at the origin. At this point if the knot with two horizontal edges is not planar, we scale it up until all its edges other than the two leaving the critical vertex are far from the other two knots, but keeping the vertex at the origin fixed - the scaling is unnecessary if this knot is planar.

The knots and disks are not in general position because moving them less than $\epsilon$ can eliminate or increase intersections. We address this by moving the two knots minimally to put them in general position with a specified intersection pattern as follows.

We translate the knot with its maximum at the origin up a bit and the knot with its minimum at the origin is pushed down slightly as in Figure 3. Finally rotating one of the knots a tiny bit and translating another is enough to create the following intersection pattern: each $D_r$ will intersect $D_s \cup D_t$. 

3
\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{The Knots are initially moved so they have an extreme vertex at the origin. \(\{e_1, f_1\} \subset K_1\) lies in the \(xy\)-plane, \(\{e_2, f_2\} \subset K_2\) and \(\{e_3, f_3\} \subset K_3\) lies in the \(yz\)-plane. \(K_1\) has its global extremum at the origin with respect to some vector in the \(xy\)-plane. \(K_2\) has its global maximum with respect to \(z\) at the origin and \(K_3\) has its global minimum with respect to \(z\) a the origin. On the right we see a vector \(\vec{w}\) with its head at \(v_2\) and its tail at the other vertex of \(f_2\) and we see the vector \(\vec{v}\) with its head on \(v_3\) at the origin and its tail between \(e_3\) and \(f_3\). Although it is vertical here, it need not be in general.}
\end{figure}

\(r \neq s \neq t\), in a pair of crossed arcs, one of which will have its end points on \(K_r\) and the other having its end points in the interior of \(D_r\) as in Figure 8. This intersection pattern occurring on all three disks is well known to imply the link is the Borromean Rings.

With the outline in mind, we now provide the complete details of the proof.

\textit{Proof of Theorems 2.1 and 2.2}: The two proofs are nearly identical, so it will be easy to prove both at the same time. We start by arguing that we may use rigid transformations of \(\mathbb{R}^3\) to position \(K_1\), \(K_2\), and \(K_3\) as they appear in Figure 2 and then use a translation to arrive at Figure 3. In the case where one of the components is planar, let it without loss of generality, be \(K_1\). For each \(K_i\) we pick an extremal vertex (a vertex that is a global maximum with respect to some direction vector) which we will call \(v_i\). The edges adjacent to \(v_i\) will be called \(e_i\) and \(f_i\).

We initially position \(v_1\), \(v_2\), and \(v_3\) at the origin. To be specific, for \(K_1\) place \(e_1\) in the \(xy\)-plane so that \(v_1\) is at the origin and \(e_1\) lies on the
Figure 3: We shift $K_2$ slightly in the direction of the vector $\vec{w}$ and $K_3$ slightly in the direction of $\vec{v}$.

(negative) $y$-axis. Fixing this edge rotate $K_1$ until $f_1$ lies in the $xy$-plane and has positive values for $x$ coordinates aside from at the point $v_1$ which has $x$ coordinate 0. Place $K_2$ so that $v_2$ is at the origin, $e_2$ and $f_2$ lie in the $yz$-plane, and so that $v_2$ is the unique global maximum of $K_2$ with respect to $z$. Place $K_3$ so that $v_3$ is at the origin, $e_3$ and $f_3$ lie in the $yz$-plane, but so that $v_3$ is the unique global minimum of $K_3$ with respect to $z$ (or equivalently it is the unique global maximum with respect to $-z$). If any two edges of $\{e_2, f_2, e_3, f_3\}$ are co-linear then rotate $K_2$ slightly around the $x$ axis. A sufficiently small rotation will fix $v_2$ and preserve the required properties above.

The proof will be dependent on $f_2$ being the least steep edge from the collection $\{e_2, f_2, e_3, f_3\}$ (in other words the absolute value of the slope of $f_2$ in the $yz$-plane is less than the absolute value of the slopes of $e_2$, $e_3$ and $f_3$ in the plane). This can easily be achieved through rotations and possible relabeling of $K_2$ and $K_3$ (possible moves include rotating $K_2$ or $K_3$ 180 degrees around the $z$ axis and switching the labels of the corresponding pair $e_i$ and $f_i$ so that the edge labeled $e_i$ remains to the left of the edge labeled $f_i$ in the $yz$ plane, and/or possibly rotating $K_2 \cup K_3$ 180 degrees around the $x$ axis and
switching the labels of the two knots as well as the names of the edges).

Lemma 2.3. If \( v_i \) is a unique global maximum for an unknot \( K_i \) (with respect to some direction vector) then we may choose a disk \( D_i \) whose boundary is \( K_i \) and which also has the point \( v_i \) as its unique global maximum in the same direction.

Proof. This is done with a standard argument in three manifold topology called an innermost loop argument. Since there is a plane that intersects \( K_i \) in \( v_i \) and otherwise contains \( K_i \) entirely on one side of it, we can also find a sphere tangent to the plane at \( v_i \), intersecting the knot only in \( v_i \) and which otherwise contains \( K_i \) entirely inside of it (as the radius of the spheres tangent to the plane at \( v_i \) goes to infinity, the spheres limit on the plane). Now pick an embedded disk \( D_i \) for \( K_i \) whose interior intersects \( S \) transversally in a minimal number of components. Since \( K_i \cap S \) is a single point there are no arcs of intersection in \( S \cap D_i \). This means all remaining intersections may be assumed to be circles (if the surfaces intersect transversally away from \( v_i \) this will be true. If not we may move the disk an arbitrarily small amount ensuring that they do intersect transversally). If the set of circles is nontrivial take an innermost circle on \( S \) (one of the components of \( S \cap D_i \) that bounds a disk on \( S \) disjoint on its interior from \( S \cap D_i \)) and cut and paste \( D_i \) replacing the component of \( D_i \) not intersecting \( S \) that is bounded by this circle and does not contain \( K_i \) by the corresponding subset of \( S \). Pushing the new disk slightly off of \( S \) gives a new disk \( D_i' \) that intersects \( S \) fewer times than \( D_i \) did yielding a contradiction to the existence of circles of intersection and showing that we may assume \( D_i \cap S = v_i \).

\[ \square \]

When positioning and scaling \( D_1 \), we will be interested in the distance between two points \( p \) and \( q \) on \( D_1 \) measured in two different ways. The first will be \( \rho(p, q) \), the length of the shortest path on \( D_1 \) between \( p \) and \( q \). The second will be \( d(p, q) \), the distance between \( p \) and \( q \) in \( \mathbb{R}^3 \). Obviously since \( D_1 \) is embedded in \( \mathbb{R}^3 \) \( d(p, q) \leq \rho(p, q) \). We want to argue that we may pick \( D_1 \) such that for \( v_1 \) together with some sufficiently small \( \epsilon > 0 \) the points on \( D_1 \) less than \( \epsilon \) from \( v_1 \) will be exactly the same set of points whether measured in the \( \rho \) metric or the \( d \) metric.

We may assume that there is some \( \epsilon_1 > 0 \) such that the portion of \( D_1 \) less than \( \epsilon_1 \) away from \( v_1 \) in the \( \rho \) metric is a subset of the flat triangle contained in the \( xy \) plane running between \( e_1 \) and \( f_1 \) (if \( K_1 \) is planar we will just choose
$D_1$ to be the planar subset of the $xy$ plane bounded by $K_1$). Now let $J_1$ be $D_1$ after deleting an open set consisting of the points in $D_1$ less than $\epsilon_1$ away from $v_1$ using the $\rho$ metric. $J_1$ is compact and does not contain $v_1$. As a result, we now can find a positive number that is the minimum distance from $v_1$ to any point in $J_1$ using metric $d$. Choose a positive number less than the minimum of this value and $\epsilon_1$ and call it $\epsilon$. Now we know that an $\epsilon$ ball around the origin in $\mathbb{R}^3$ (in the tradition metric $d$) intersects $D_1$ only in points contained in the $xy$ plane and thus the set of all points of $D_1$ less than $\epsilon$ away from $v_1$ in the $d$ metric is identical to the set of all points of $D_1$ less than $\epsilon$ away from $v_1$ in the $\rho$ metric. An analogous argument may be used for $D_2$ and $D_3$.

**Lemma 2.4.** If $K_1$ is planar then we may assume that $D_i \cap D_j, i, j \in \{1, 2, 3\}, i \neq j$ is $v_i = v_j$ the two vertices at the origin. If $K_1$ is not planar, we may scale $K_1$ up until $D_i \cap D_j, i, j \in \{1, 2, 3\}, i \neq j$ is again $v_i = v_j$.

**Proof.** Pick $D_2$ and $D_3$ to be flat near the origin and according to Lemma 2.3 so that they intersect the $xy$ plane only at the origin. This ensures that they intersect each other only at the origin. If $K_1$ is planar, the result follows by picking $D_1$ to be the flat disk totally contained in the $xy$ plane. If not then for each $i$, let $r_i$ be the maximum distance from the origin to any point of $D_i$ (in the metric $d$). Without loss of generality let $r_2 \geq r_3$. Scale $K_1$ up by multiplying by the three by three matrix $\lambda I$, where $\lambda > \frac{r_2}{r_3}$. This will ensure that any point on $D_1$ that is not in the $xy$ plane is farther from the origin than any point in $D_2$ or $D_3$. Thus the only place where $D_2$ or $D_3$ could intersect $D_1$ would be where they intersect the $xy$ plane, which is only at the origin.

This is the only scaling we need in the proof of Theorem 2.1 and no scaling is needed in the proof of Theorem 2.2. Otherwise the proofs of the two theorems are identical. All other transformations will be translations and rotations. We now fix $K_1$ for the rest of the proof and move the other two knots slightly starting with $K_2$.

From now on when we talk about distance we will only refer to the standard metric $d$ in $\mathbb{R}^3$ and not the metric $\rho$.

**Lemma 2.5.** Given disks $\{D_1, D_2, D_3\}$ intersecting only at the origin and bounded by knots $\{K_1, K_2, K_3\}$ as above, then for any $\epsilon > 0$ if we let $\{J_1, J_2, J_3\}$
be equal to \( \{D_1, D_2, D_3\} \) minus the portion of the \( D_i \)'s in an open \( \epsilon \) ball around the origin, and given any translation of \( \mathbb{R}^3 \) acting on a given \( D_i \) or any rotation of that \( D_i \) about a fixed axis \( l \), then there exists an \( \epsilon' \) such that any translation of distance less than \( \epsilon' \) or rotation about \( l \) of angle less than \( \epsilon' \) leaves the \( J_i \)'s pairwise disjoint.

Proof. Since \( J_i \cap J_j = \emptyset \) and they are both compact, there is a positive minimal distance \( s \) between any point in \( J_i \) and \( J_j \). Setting \( \epsilon' < s \) will ensure that translating one of the disks less than \( \epsilon' \) cannot create an intersection. Similarly after fixing an axis of rotation we can pick a small enough angle such that rotating \( J_i \) will move no point of \( J_i \) more than \( \epsilon' \).

The virtue of Lemma 2.5 is that we now know that during all our remaining manipulations of the disks and knots no new intersections will be introduced and all intersections will occur in an \( \epsilon \) ball neighborhood of the origin on the flat triangular pieces of the \( D_i \)'s running between \( e_i \) and \( f_i \) (ie the portions of the \( D_i \)'s not contained in the \( J_i \)'s).

Let \( \vec{w} \) be the vector with its head at \( v_2 \) and parallel to \( f_2 \) (its tail may be thought of as lying on the other vertex of \( f_2 \)) as in Figure 2. Translate \( K_2 \) by adding \( \epsilon \ast \vec{w} \) to every point on \( K_2 \) for a sufficiently small \( \epsilon \) in order to translate \( K_2 \) (and \( D_2 \)) minimally up in a direction parallel to \( f_2 \). We want to be certain that none of the vertices of \( K_2 \) other than \( v_2 \) rise up to or above the \( xy \)-plane, that \( e_2 \cap e_1 \) remains nontrivial, and that \( v_3 \) is the only critical point of \( K_3 \) that \( v_2 \) rises above. Choosing a sufficiently small \( \epsilon \) will ensure all of these properties, as would any positive translation smaller than \( \epsilon \). Now \( v_2 \) is very close to, but above \( v_1 \), \( f_2 \) intersects the origin (\( v_1 \)), and \( e_2 \) intersects \( e_1 \) in some point other than \( v_1 \). \( K_1 \) and \( K_2 \) look as they do in Figure 3 and we need to reposition \( K_3 \) to match the figure.

With the exception of \( v_2 \) and \( v_3 \) all critical points of \( K_3 \) are now above all critical points of \( K_2 \). The fact that \( f_2 \) is not as steep as \( e_3 \) and \( f_3 \) ensures that both \( e_3 \) and \( f_3 \) are on the same side of the line containing \( f_2 \) in the \( yz \)-plane. \( f_2 \) being less steep than \( e_2 \) and on its right in the \( yz \) plane ensures that \( f_2 \) has negative slope. Let \( \vec{v} \) be a vector with its head on \( v_3 \) at the origin and its tail between \( e_3 \) and \( f_3 \) as in Figure 2. Translate \( K_3 \) by \( \delta \ast \vec{v} \) for a small \( \delta \). This keeps \( e_3 \) and \( f_3 \) in the \( yz \)-plane. Choosing a sufficiently small \( \delta \) again makes sure that the critical points of \( K_3 \) aside from \( v_3 \) remain above the critical points of \( K_2 \), all the critical points of \( K_2 \) other than \( v_2 \) remain below all the critical points of \( K_3 \), and that \( K_2 \cap K_3 \) consists of exactly two
points, $e_3 \cap f_2$ and $f_3 \cap f_2$. Finally to complete the figure pick $l$, a line in
the $yz$-plane parallel to $f_2$, but separating $f_2$ from $v_3$. Let $l \cap e_3$ be called
$p_e$ and let $l \cap f_3$ be called $p_f$. We must pick $l$ close enough to $f_2$ so that $p_e$
is above $e_1$. This is easy to do since $e_1$ lies on the $y$-axis and we need only
make sure that $p_e$ has positive $z$-coordinate. The fact that $f_2$ goes through
the origin, has negative slope, and that the point $e_3 \cap f_2$ is to the left of the $z$
axis ensures that it has positive $z$ coordinate so if $l$ is sufficiently close to $f_2$
then $p_e$ will, too. Since $f_2 \cap f_3$ occurs to the right of the origin, it has negative
$z$-coordinate, and $p_f$ is below this point, $p_f$ will have negative $z$-coordinate.

Recall that we have already established that in Figure 3 we may assume
the $D_i$ are all disjoint from the neighborhood of the origin depicted except
in the obvious flat triangular subdisks and that the $D_i$ are disjoint from each
other outside of the figure.

Now we want to put the knots and disks in general position. This process
will take us from Figure 3 to Figure 6. Because general position is always
easy to attain with infinitesimally small transformations we can pick a small
number $\epsilon$ and no point of the knots or disks will move more than $\epsilon$ over
the rest of the proof. This ensures that the only new intersection patterns
between the disks will be the result of local changes in the current intersection
patterns.

For the rest of the proof $K_1$ will remain fixed. Now rotate $K_3$ around $l$
so that the $x$ coordinate of $v_3$ becomes negative and so that $D_3$ is in general
position with respect to both $D_1$ and $D_2$ (although $D_1$ and $D_2$ are still not
in general position with respect to each other). Rotating by a small enough
angle will ensure that no point on $K_3$ or $D_3$ moves more than $\epsilon$. The rotation
will fix $p_e$ and $p_f$, will cause all the points on the same side of $l$ as $v_3$ to have
negative $x$ coordinates and all the points of $e_3 \cup f_3$ on the other side to have
positive $x$ coordinates. Before rotating $D_3$ intersected $D_1$ in a single arc,
subset of $e_1$ running from $e_3 \cap e_1$ to $v_1$, including the single point $l \cap D_1$.
After the rotation $D_1 \cap D_3$ will remain an arc, $l \cap D_1$ will be one endpoint,
and the arc of intersection will rotate about this point and the other end
point will move away from the origin ($v_1$) to a point on $f_1$ with positive $x$
coordinate as in Figure 7.

Before rotating $D_3$, $D_2 \cap D_3$ was a triangle formed by intersecting
the triangle subset of $D_2$ running from $e_2$ to $f_2$ and the analogous triangle on
$D_3$ from $e_3$ to $f_3$. The two triangles and thus the intersection contained the
portion of $l$ running from $p_e$ to $p_f$. After rotating this portion of $l$ will be
the only portion of $D_3$ near the origin contained in the $xy$ plane. Since we

9
Figure 4: The initial intersections of the disks $D_1 \cap D_2$, $D_1 \cap D_3$, and $D_2 \cap D_3$ respectively are shown in gold. They are not yet in general position.

have already established that all intersections will occur near the origin this means that $D_2 \cap D_3$ is exactly the arc of $l$ running from $p_e$ to $p_f$. Now $D_3$ is in general position with respect to both $D_1$ and $D_2$.

Finally we must translate $K_2$ and $D_2$ slightly so that $D_1 \cap D_2$ is in general position. This will move $D_2 \cap D_3$ infinitesimally, but since they are already in general position and the move will be minimal it will not be enough to change the intersection pattern of those two topologically so for our purposes we may think of it as essentially unchanged. Before translating $D_2 \cap D_1$ is the subset of the edge $e_1$ running from $e_1 \cap e_2$ to $v_1 = e_1 \cap f_2$. Let $\vec{u}$ be a vector in the $xy$-plane with tail at the origin ($v_1$) and head on the triangular portion of $D_1$ between $e_1$ and $f_1$ as in Figure 5. Translate $K_2$ by $\rho \ast \vec{u}$ where
Figure 5: $K_3$ has been rotated and the Borromean rings will be formed once we translate $K_2$ slightly in the direction of horizontal vector $\vec{u}$ which lies in the $xy$-plane with its tail at the origin and its head between $e_1$ and $f_1$.

$|\rho \ast \vec{u}|$ is small enough to satisfy Lemma 2.5. Since $D_2$ and $D_3$ were in general position and our translation was minimal, $D_2 \cap D_3$ remains an arc as before (although it is no longer a subset of $l$). $D_1$ and $D_2$ now are in general position and $D_1 \cap D_2$ becomes an arc from $e_2 \cap D_1$ to $f_2 \cap D_1$ that is parallel to, but now disjoint from $e_1$.

All the disks are now in general position and the link looks locally like Figure 6. The disks now each intersect the union of the other two in a cross as in Figure 8. It is not hard to show that this intersection pattern can only result from the Borromean Rings. See, for example, [H2].

## 3 Conjectures and open questions

We conclude with some open questions and conjectures.

**Conjecture 3.1.** Any three planar curves can be used to form the Borromean rings as long as at least one is not a circle.
Planar was convenient and was necessary at times for the proofs in [H3], but it is not clear that the theorem fails without it even if this proof does.

**Conjecture 3.2.** Any three unknots can be used to form the Borromean rings through rigid transformations and scaling applied to the individual components as long as at least one is not a circle.

**Question 3.3.** Can any three curves be used to form the Borromean rings through rigid transformations applied to the individual components as long as one is not a circle? (Here scaling is not allowed.)
Figure 7: The Intersections of the disks in the final format $D_1 \cap D_2$, $D_1 \cap D_3$, and $D_2 \cap D_3$ respectively are shown in black. Note that each time the black arc has end points on one of the knots (in the top left $K_1$, top right $K_2$, and bottom $K_3$) and on the interior of the disk bounded by the other knot. If all three arcs were drawn in the same picture we would see that the top two form a cross in the $xy$-plane intersecting in a single point on the interior of both arcs. The third arc intersects the $xy$-plane they would cross in a single point on its interior, which also is the unique point where it intersects the other two arcs.
Figure 8: $D_r \cap (D_s \cup D_t)$ looks like the figure above for any distinct choices of $r, s, t \in \{1, 2, 3\}$. This can only happen in the case of the Borromean rings.

4 References


H3 H. N. Howards, Forming the Borromean Rings from planar curves. (Preprint).