Andrew James Critch

Resolving the Banach-Tarski Paradox: Inseparability of Rigid Bodies

B.Sc. Honours Dissertation,
Memorial University of Newfoundland,
April, 2006

Abstract (and Introduction)

The unified purpose of this paper is to *prove* the Banach-Tarksi *Theorem* (BTT) and to *resolve* the Banach-Tarksi *Paradox* (BTP). The BTT is a rigorous mathematical result which is usually interpreted as saying that a (solid) unit ball can be ″taken apart″ into a finite number of rigid pieces which can then be rigidly ″put back together″ to form two unit balls. This interpretation is called the Banach-Tarksi Paradox, because it contradicts our usual expectations about the conservation of mass/measure.[1] To ″resolve″ the Paradox will mean to find a more intuitively acceptable interpretation of the BTT and to explain what is ″wrong″ with the paradoxical interpretation.

The BTT is interpreted as ″taking apart″ a ball by applying isometries to pieces (subsets) of the ball. However, these isometries account only for the initial and final positions of the pieces, and do not describe their intermediate positions (i.e., their trajectories). Consider how, when restricted to two dimensions, it is impossible to ″take apart″ the pieces of a jigsaw puzzle, even though isometries can be used to map the pieces of an assembled puzzle to those of a disassembled one. This is because the pieces of a jigsaw puzzle are ″interlocked″ (later, ″*inseparable*″). It is therefore conceivable that, when restricted to three dimensions, it is *not* possible to ″take apart″ a unit ball into a finite number of pieces and ″put them back together″ to make two unit balls.

This is precisely the motivation for this paper, which shows how BTT proofs (at least those known to this author) make use of pieces which are indeed inseparable like those of a jigsaw puzzle. In fact, unlike the pieces of a jigsaw puzzle, which can be easily ″cut″ into smaller pieces that *can* be taken apart in 2*D*, the BTT pieces are interlocked in a more complicated way that is not so easily remedied in 3*D*.

[1] The BTT is often *called* the Banach-Tarksi Paradox, but this author chooses to distinguish between the Theorem itself and it's confusing intuitive interpretation.

Contents

0 Preliminaries.

The reader is encouraged to skim this section or skip it completely, referring to it later as needed.

Throughout this paper, an attempt will be made to give definitions in sufficiently general terms as to be useful elsewhere, but sufficiently specific terms as to be easily understandable in context.

Some Common Conventions dealing with functions and relations:

- If R is a relation (i.e. a set of ordered pairs), the *domain* and *range* of R are respectively

$$\operatorname{Dom} R := \{x \mid \exists y,\ \ (x,y) \in R\}, \qquad \operatorname{Rng} R := \{y \mid \exists x,\ \ (x,y) \in R\}$$

- For a set S (usually a subset of $\operatorname{Dom} R$), the *point-wise image* of S under R is defined as
$$R[S] := \{y \mid \exists s \in S,\ \ (s,y) \in R\}.$$
For $x \in \operatorname{Dom} R$, if no confusion will result, $R[x]$ will denote $R[\{x\}]$. In this way, R can be regarded as a "multivalued function" from $\operatorname{Dom} R \to \operatorname{Rng} R$.

- For ease of reference, $R : X \to Y$ will indicate that $\operatorname{Dom} R = X$ and $\operatorname{Rng} R \subseteq Y$.

In this paper, it should be emphasized that a relation is considered *solely* as a set of ordered pairs, and so we may write $R : X \to Y$ and $R : X \to Z$ without referring to two different relations: the association of R with a so-called "target" set is purely extrinsic. As well, a function is merely a special case of a relation, and so our discussions of relations also apply to functions.[2]

- The *inverse* of R is the set $\{(y,x) | (x,y) \in R\}$, which does not in general "cancel" the action of R like an algebraic inverse. Since any function f is a relation, f^{-1} exists as a relation, but is a function iff f is injective.

- The *restriction of* R: $X \to Y$ to a set S is the relation $\{(s,y) \in R \mid s \in S\}$ denoted by $R|_S$. When restricting a function $f : X \to Y$, usually $S \subseteq X$ so that $f|_S$ is the function $S \to Y$ given by $f|_S(s) = f(s)$.

- The *union* of two relations $Q \cup R$ is simply their set theoretic union as sets of ordered pairs. For functions $f : A \to B$ and $g : C \to D$, the union $f \cup g$ is a function iff $f|_{A \cap B} = g|_{A \cap B}$.

To illustrate a use of the above terminology, consider the real function sin. We can say that $\sin : \mathbb{R} \to \mathbb{R}$ and equally well that $\sin : \mathbb{R} \to [-1,1]$, since it is true that $\sin[\mathbb{R}] = [-1,1]$. Note that $\sin^{-1} : [-1,1] \to \mathbb{R}$ is a relation but not a function, and for instance $\sin^{-1}[0] = \pi\mathbb{Z}$. However, since $\sin|_{[-\pi,\pi]}$ is injective,

[2] Henceforth, R will always denote a relation, and f will always denote a function.

it has an inverse function, which is usually called $\arcsin : [-1,1] \to [-\pi,\pi]$. Finally, as an example of a function union, observe that if we define the real functions $f : \mathbb{R} \to \mathbb{R}$, $g : [0,\infty] \to \mathbb{R}$, and $h : [-\infty,0] \to \mathbb{R}$ by $f(x) = x|x|$, $g(x) = x^2$, and $h(x) = -x^2$, then it is true that $f = g \cup h$.

Some Nonstandard Conventions which are adopted for the particular convenience of this paper in dealing with collections of sets (although some of them can be very convenient in other settings as well):

- If $\mathcal{C}$ is a collection of subsets of X (i.e. $\mathcal{C} \subseteq \boldsymbol{\mathcal{P}}(X)$, the power set of X), the *piece-wise image* of $\mathcal{C}$ under R is

$$R\{\mathcal{C}\} := \{R[S] : S \in \mathcal{C}\}.$$

- A *partition* $\mathcal{P}$ is a collection of non-empty, disjoint sets. To be more specific, $\mathcal{P}$ is said to be a partition *of* X in the case that $\bigcup \mathcal{P} = X$.

- The *restriction of* $\mathcal{P}$ *to* $S \subseteq X$ is the partition

$$\mathcal{P}|_S := \{P \cap S \neq \emptyset : P \in \mathcal{P}\}.$$

- The *conjunction* of two partitions $\mathcal{P}$ and $\mathcal{Q}$, is the set

$$\mathcal{P} \wedge \mathcal{Q} := \{P \cap Q \neq \emptyset : \ P \in \mathcal{P}, Q \in \mathcal{Q}\}.$$

From the above, for example, it follows directly that if $f : X \to Y$ and $\mathcal{P}$ is a partition of Y then $f^{-1}\{\mathcal{P}\}$ is a partition of X (with the convention that we eliminate the empty set from a collection if it is to be considered as a partition). The terminology used can be motivated by the following discussion:

A function f with $\operatorname{Dom} f = X$ defines an equivalence relation $\overset{\mathcal{P}}{\sim}$ given by $x \overset{\mathcal{P}}{\sim} y \Leftrightarrow f(x) = f(y)$, which in turn corresponds to a unique partition $\mathcal{P}$ of X. For $S \subseteq X$, the restricted partition $\mathcal{P}|_S$ is then the partition of S thus induced by the restricted function $f|_S$. Also, the conjunction $\mathcal{P} \wedge \mathcal{Q}$ is that partition corresponding to the conjunction of the equivalences "$x \overset{\mathcal{P}}{\sim} y \ \wedge \ x \overset{\mathcal{Q}}{\sim} y$."

1 Arrangements.

This section defines the concept of an "arrangement" to agree with the everyday notion that a solved jigsaw puzzle is a **re**arrangement of an unsolved one (the author has found the prefix "**re**" to be superfluous and it will henceforth be omitted). Some terminology is created in tandem with the definition to facilitate the future development of results about arrangements.

The concept of an arrangement effectively replaces the notion of "decomposition" which has previously been used to express the BTT. Arrangements have the advantage of being naturally more convenient and understandable in proofs, because they can be composed like mappings. In fact, arrangements can be viewed as morphisms in a category whose objects are rigid bodies (metric subspaces Euclidean space). This paper presupposes no knowledge of category theory, and does not spend time developing this category-theoretic interpretation.

Definition (1.1)**.** *(Arrangements)*

a) For $A \subseteq \mathbb{R}^n$, a *k-arrangement* of A is defined to be a pair $(\phi, \mathcal{P})$ where ϕ is a map $A \to \mathbb{R}^n$, $\mathcal{P}$ is a partition of A with $|\mathcal{P}| \leq k$, $\phi\{\mathcal{P}\}$ partitions $\phi[A]$, and $\forall\, P \in \mathcal{P}$, $\phi|_P$ is an isometry. (note this implies ϕ is an injection)

b) $(\phi, \mathcal{P})$ is called *finite* if $|\mathcal{P}|$ is finite. In this paper, all arrangements are assumed to be finite.

c) $(\phi, \mathcal{P})$ is said to be *onto* B if $\phi[A] = B$. We write $(\phi, \mathcal{P}) : A \overset{k}{\simeq} B$. Note that, in this case, $(\phi^{-1}, \phi\{P\}) : B \overset{k}{\simeq} A$, so $\overset{k}{\simeq}$ is a symetric (and clearly reflexive) relation. Thus, A and B are said to be "k-equiarrangeable".

d) If we wish to claim only the *existence* of an appropriate $(\phi, \mathcal{P})$ and k, we may choose to omit unwanted references by writing $A \underset{\phi}{\overset{k}{\simeq}} B$, $A \overset{k}{\simeq} B$ or simply $A \simeq B$.

e) If $(\phi, \mathcal{P}) : A \overset{k}{\simeq} B'$ and $B' \subseteq B$, we write $(\phi, \mathcal{P}) : A \overset{k}{\preceq} B$ and analogous expressions. In this case, $(\phi, \mathcal{P})$ is said to be *into* B, and A is said to be k-arrangeable into B.

It should be noted in (a) that the number k sets an *upper bound* on $|\mathcal{P}|$, the number of "pieces" required for the arrangement, and does not specify the exact number of pieces. This policy is adopted since most theorems will set upper bounds on the number of pieces rather than determine the exact number. If interest warrants, a k-arrangement can be called *exact* if $|\mathcal{P}|$ is *exactly* k. (Alternatively, one could always interpret the number of pieces to be exactly k and allow a number of "empty pieces" if the number if nonempty pieces is $< k$. A similar approach to this will be used in §6.)

Also, note that in (d), ommiting a reference to a particular ϕ or k results in a weaker and therefore *different* relation. This is important since, as we shall see, the relation $\simeq$ will exhibit an important property that $\overset{k}{\simeq}$ does not.

We now introduce some operations on arrangements because of the results that easily follow from their definitions (once the operations are identified, the proofs are routine set theory and have therefore been omitted).

Definitions/Theorems (1.2). *(Useful operations and results)*

a) If $A \subseteq A'$ and $(\phi, \mathcal{P}) : A' \overset{k}{\simeq} B$, we define the *restriction of* $(\phi, \mathcal{P})$ *to* A as the k-arrangement

$$(\phi, \mathcal{P})|_A := (\phi|_A, \mathcal{P}|_A).$$

In this case, $(\phi, \mathcal{P})|_A : A \overset{k}{\preceq} B$, so in general,

$$A \subseteq A' \overset{k}{\simeq} B \;\Rightarrow\; A \overset{k}{\preceq} B.$$

b) For $A \cap C = \emptyset = B \cap D$, if $(\phi, \mathcal{P}) : A \overset{j}{\simeq} B$ and $(\psi, \mathcal{Q}) : C \overset{k}{\simeq} D$, we define their *disjoint union* as the $(j + k)$-arrangement

$$(\phi \cup \psi, \mathcal{P} \cup \mathcal{Q}) : A \cup C \overset{j+k}{\simeq} B \cup D.$$

Notice that if $A \cap C \neq \emptyset$ we may replace C by $C \setminus A$ and $(\psi, \mathcal{Q})$ by $(\psi, \mathcal{Q})|_{C \setminus A}$ to achieve the above conditions. Hence, only $B \cap D = \emptyset$ is required for the implication

$$A \overset{j}{\simeq} B \text{ and } C \overset{k}{\simeq} D \;\Rightarrow\; A \cup C \overset{j+k}{\simeq} B \cup D.$$

c) If $(\phi, \mathcal{P}) : A \overset{j}{\simeq} B$ and $(\psi, Q) : B \overset{k}{\simeq} C$, we defined their *composition* as the jk-arrangement

$$(\psi, \mathcal{Q}) \circ (\phi, \mathcal{P}) := (\psi \circ \phi, \mathcal{P} \wedge \psi^{-1}\{Q\}).$$

In this case, $(\psi, \mathcal{Q}) \circ (\phi, \mathcal{P}) : A \overset{jk}{\simeq} C$, so in general,

$$A \overset{j}{\simeq} B \overset{k}{\simeq} C \;\Rightarrow\; A \overset{jk}{\simeq} C.$$

d) *All of the above results immediately hold when* $\simeq$ *is replaced by* $\preceq$

Results (c) and (d) show that the $\simeq$ and $\preceq$ relations (which do not discriminate according to a fixed value of k, but only require that each partition be finite) are transitive, and in particular that $\simeq$ is an equivalence relation.

The reader is now invited to consider whether $A \preceq B \preceq A \Rightarrow A \simeq B$. That is, suppose one can arrange A to form a part of B and arrange B to form a part of A. Is it then possible, still using only a finite number of pieces, to arrange A to form *precisely* B, with no leftovers or overlaps? The question may seem particularly daunting in the cases of infinite, uncountable, unmeasurable, and otherwise pathological sets. The converse is clearly true, but those who briefly attempt and fail to prove the forward implication may be inclined to suspect the existence of a counterexample. Nevertheless, truth prevails, and the proposition is indeed true:

Theorem (1.3)**.** *(The Banach-Schröder-Bernstein Theorem)*

$$A \overset{j}{\preceq} B \overset{k}{\preceq} A \Rightarrow A \overset{j+k}{\simeq} B.$$

Proof: Let $(\phi, \mathcal{P}) : A \overset{j}{\preceq} B$ and $(\psi, \mathcal{Q}) : B \overset{k}{\preceq} A$. Define, recursively,

$$A_0 = A, \;\; A_1 = \psi[B], \;\; B_0 = B, \;\; B_1 = \phi[A]$$
$$A_{i+2} = \psi[\phi[A_i]], \;\; B_{i+2} = \phi[\psi[B_i],$$

so that $\phi[A_i] = B_{i+1}$, $\psi[B_i] = A_{i+1}$, $A_{i+1} \subseteq A_i$, and $B_{i+1} \subseteq B_i$. Next, let

$$\begin{aligned} A^0 &= \{A_0 \setminus A_1 \;\cup\; A_2 \setminus A_3 \;\cup\; A_4 \setminus A_5 \;\cup\; \cdots\} \\ A^1 &= \{A_1 \setminus A_2 \;\cup\; A_3 \setminus A_4 \;\cup\; A_5 \setminus A_6 \;\cup\; \cdots\} \\ A^\infty &= \{A_0 \;\cap\; A_1 \;\cap\; A_2 \;\cap\; A_3 \;\cap\; A_4 \;\cap\; \cdots\} \end{aligned}$$

and

$$\begin{aligned} B^0 &= \{B_0 \setminus B_1 \;\cup\; B_2 \setminus B_3 \;\cup\; B_4 \setminus B_5 \;\cup\; \cdots\} \\ B^1 &= \{B_1 \setminus B_2 \;\cup\; B_3 \setminus B_4 \;\cup\; B_5 \setminus B_6 \;\cup\; \cdots\} \\ B^\infty &= \{B_0 \;\cap\; B_1 \;\cap\; B_2 \;\cap\; B_3 \;\cap\; B_4 \;\cap\; \cdots\} \end{aligned}$$

Now, $\{A^0, A^1, A^\infty\}$ is a partition of A and $\{B^0, B^1, B^\infty\}$ is a partition of B. Furthermore, since injections preserve intersections, unions, and set differences, observe directly that

$$\phi[A^0] = B^1, \;\; \phi[A^\infty] = B^\infty, \;\; \textit{and} \;\; \psi[B^0] = A^1.$$

Hence, we have the arrangements

$$\begin{aligned} (\phi, \mathcal{P})|_{A^0 \cup A^\infty} &: \;\; A^0 \cup A^\infty \overset{k}{\simeq} B^1 \cup B^\infty \\ (\psi, \mathcal{Q})|_{B^0} &: \;\; B_0 \overset{j}{\simeq} A_1 \end{aligned}$$

and taking the appropriate union gives

$$A = A^0 \cup A^1 \cup A^\infty \overset{j+k}{\simeq} B^0 \cup B^1 \cup B^\infty = B. \quad \square$$

Thus, the $\preceq$ relation (which does not discriminate according to a fixed value of k, but requires only that each partition be finite) is antisymetric up to equiarrangeability. In other words, $\preceq$ defines a partial ordering on $\simeq$ equivalence classes. This result, which even gives a bound on the number of pieces required, can be used to design many new arrangements.

2 Rotational arrangements of spheres; Sectoids.

We will now explore various arrangement properties of the sphere and its subsets. Only rotational isometries will be used, since rotations can "happen" in "everyday space" whereas reflections cannot. As well, for *most* subsets A of the sphere (i.e., those which contain non-coplanar points), only isometries which leave the origin fixed will map the subset back into the sphere. Our results will then be naturally extended to the "solid" unit ball (by defining "sectoids") in preparation for proving the Banach-Tarski Theorem in §3.

Notation.

- $S^n := \{x \in \mathbb{R}^{n+1} : |x| = 1\}$, the unit sphere.
- Given a subset $A \subseteq S^n$, $A' := S^n \setminus A$, the "complement" of A.
- $SO(n)$ denotes the set of rotations about the origin in $\mathbb{R}^n$. Hence $SO(n+1)$ is the rotational symetry group of S^n.

Our first theorem is inspired by the scenario $\mathbb{N} \setminus \{1\} - 1 = \mathbb{N}$, where a point removed from an infinite set can be "replenished" by an isometry (in this case, the translation $x \mapsto x - 1$).

Theorem (2.1). *If $P \subseteq S^n$ is countable, then $S^n \overset{2}{\simeq} P'$.*

Lemma. We begin by showing that *there exists a rotation, $\alpha \in SO(n+1)$, such that the sets P, αP, $\alpha^2 P, \ldots$ are all disjoint.*

First observe that there are uncountably many axes of rotations in $SO(n+1)$ and only countably many axes intersect P, hence we may choose an axis disjoint from P, and let α_θ denote a rotation through an angle, θ, about that axis (so that $\forall\, p \in P$, p is not on the axis of α_θ, hence $\alpha_\theta p \neq p$).

Next, for each $p \in P$ let $A_p = \{\theta \in [0, 2\pi) : \alpha_\theta p \in P\}$, so A_p is countable, and $B_p = \{\theta \in [0, 2\pi) : \exists n > 0,\ \ \alpha_\theta^n \in A_p\}$ and B_p is also countable. Let

$$B = \bigcup_{p \in P} B_p = \{\theta \in [0, 2\pi) : \exists n > 0,\ \ P \cap \alpha_\theta^n P \neq \emptyset\}$$

and since B, the set of "bad" angles, is countable by construction, there are uncountably many $\theta \in [0, 2\pi)$ such that $\alpha_\theta \notin B$, i.e, such that P, $\alpha_\theta P$, $\alpha_\theta^2 P, \ldots$ are disjoint. Hence, fix $\alpha = \alpha_\theta$ with this property.

Using this Lemma, we now proceed with the

Proof: Let $Q = P \cup \alpha P \cup \alpha^2 P \cup \cdots$. We have $Q' \overset{1}{\simeq} Q'$ by the identity, and $Q \underset{\alpha}{\overset{1}{\simeq}} \alpha Q = \alpha P \cup \alpha^2 P \cup \cdots$, hence taking the union gives

$$S^n = Q \cup Q' \overset{2}{\simeq} \alpha Q \cup Q' = S^n \setminus P = P'. \quad \square$$

Further results on the sphere will rely on the fact that $SO(n)$ contains subgroups which are *free groups*. Free groups are algebraically very simple, but because they have so few rules describing them, they can exhibit some complex and surprising behaviour that one would not expect to find in a group of rotations in space.

Definition (2.2). *(Free group conventions)*

- A *word* on characters $a_1, \ldots, a_n$ will be distinguished as a string of (perhaps no) characters of the forms a_i and a_i^{-1} enclosed in single quotes. The default operation on words will be *concatenation* (without cancellation of inverses).
- If the characters of a word σ represent group elements, then σ_* denotes the group element "produced" by dropping quotes and considering it as a product. We define $`\,'_* = 1$.
- $\mathcal{F}\langle a_1, \ldots, a_n\rangle$ denotes the free group on characters $a_1, \ldots, a_n$.
- For $f \in \mathcal{F}\langle \alpha_1, \ldots, \alpha_n\rangle$, the unique reduced word (formal sequence of characters without adjacent inverses) σ satisfying $\sigma_* = f$ is denoted by f^*.

Essentially, f^* is merely f expressed in terms of the generators a_i and enclosed in quotes. By working with *words* f^* and the operation of *concatenation* (without cancellation), we may explore properties of the elements of F that would have been meaningless if expressed in terms of the original group operation.

- In $F = \mathcal{F}\langle a_1, \ldots, a_n\rangle$, the set of elements "beginning with f" is defined as $\omega(f) := \{x \in F : \exists y \in F,\ \ x^* = f^* y^*\}$.

The sets $\omega(f)$ are quite interesting in that they allow free groups to be "rearranged" (as sets) in peculiar but simple ways. For example, in $F = \mathcal{F}\langle\alpha, \beta\rangle$, we have

$$\alpha\left(\omega(\alpha) \cup \omega(\beta) \cup \omega(\beta^{-1}) \cup \{e\}\right) = \alpha(F \setminus \omega(\alpha^{-1})) = \omega(\alpha).$$

Consequently, the following theorem can be used to produce some strange results in $\mathbb{R}^n$:

Theorem (2.3). *$SO(n)$ contains a free group of rank 2 when $n \geq 3$.*

Proof: Let α and β be rotations about the z and x axes, respectively, each through an angle of $\cos^{-1}(\frac{1}{3})$. Then,

$$\alpha^{\pm 1} = \frac{1}{3}\begin{pmatrix} 1 & \mp 2\sqrt{2} & 0 \\ \pm 2\sqrt{2} & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix} \text{ and } \beta^{\pm 1} = \frac{1}{3}\begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & \mp 2\sqrt{2} \\ 0 & \pm 2\sqrt{2} & 1 \end{pmatrix}.$$

It can be shown by induction on product length (see, for example, "The Banach Tarksi Paradox" by So and So) that if ω is a reduced word product on

α, β, then the image $\omega(1,0,0) = \dfrac{(a, b\sqrt{2}, c)}{3^k}$ for some $a, b, c, k \in \mathbb{Z}$ with $3 \not| \, b$. In particular, $b \neq 0$ so $\omega(1,0,0) \neq (1,0,0)$, hence $\omega \neq e$. It follows that $f = \mathcal{F}\langle\alpha, \beta\rangle$ is a free group of rank 2 in $SO(3) \leq SO(n)$. □

We will now proceed to extend the above results to balls, employing the following

Notation.

- $\mathbb{B}^n := \{|x| \in \mathbb{R}^n : |x| < 1\}$
- $\overline{\mathbb{B}}^n := \{|x| \in \mathbb{R}^n : |x| \leq 1\} = \mathbb{B}^n \cup S^{n-1}$
- O denotes the origin
- Translates of these figures will be denoted by subscripts.

  e.g. $\overline{\mathbb{B}}^n_i = \mathbb{B}^n_i \cup S^{n-1}_i$ is centered at O_i (whose location may go unspecified).

To extend our rotational arrangements of spheres to arrangements of balls, all that is necessary is to identify each point p on a sphere to the set of points between p and the centre of the sphere (including p, excluding the centre). All of these points will move "in synchrony" with p under the action of central rotations.

Definition (2.5). *(Radioids and Sectoids)*

- Given a point p on a sphere S^n_i (which should be clear from context), the "radioid" from p will be defined as

$$\widehat{p} := \{\lambda p + (1-\lambda)O_i : 0 < \lambda \leq 1\} = \overline{O_i p} \setminus \{O_i\},$$

  the radial segment $O_i p$ with O_i removed.

- Given a set $A \subseteq S^n_i$, we define the *sectoid* from A as

$$\widehat{A} := \bigcup_{a \in A} \widehat{a} = \{\lambda a + (1-\lambda)O_i : 0 < \lambda \leq 1, a \in A\}.$$

The reason for assigning a terminology here is that, traditionally, proofs of the Banach-Tarski Theorem always divide the ball into these "sectoid" pieces (because they are easy to deal with, not because they *must* be used). Notice the following results (which also apply to translates):

Theorem (2.4). *For $p, q \in S^n$, $A, B \subseteq S^n$, and $\alpha \in SO(3)$,*

- $\widehat{p} \cap \widehat{q} \neq \emptyset \Leftrightarrow \widehat{p} = \widehat{q} \Leftrightarrow p = q$
- $\widehat{A} \cap \widehat{B} = \widehat{A \cap B}$, $\widehat{A} \cup \widehat{B} = \widehat{A \cup B}$, *and* $\widehat{A} \setminus \widehat{B} = \widehat{A \setminus B}$

- $\alpha\widehat{p} = \widehat{\alpha p}$, $\alpha\widehat{A} = \widehat{\alpha A}$

Since turning subsets of the sphere into sectoids will preserve set theoretic operations and the action of rotations, our previous results in S^n now have analogues in $\widehat{S^n} = \overline{\mathbb{B}}^{n+1} \setminus O$. In particular,

Corollary (2.5)**.** *If $P \subseteq S^n$ is countable, then $\widehat{S^n} \overset{2}{\simeq} \widehat{P}$.*

Proof: Apply $\widehat{\text{hats}}$ to the proof of theorem 2.1. (All the required relations will still hold when the sphere subsets are extended to sectoids).

3 Proving the BTT; Identifying the Banach-Tarksi Paradox (BTP).

We are now ready to consider the following

Theorem (3.1)**.** *The Banach-Tarski Theorem (BTT): The unit ball can be arranged using a finite number of pieces to form two solid balls of the same size and volume.*

The smallest number of pieces which permits the arrangement is 5 (i.e., 4 is known to be unattainable, and 5 has been attained). However, this paper aims not to minimize the number of pieces but to make the construction as easily understandable as possible. Hence, we will prove the following result using 10 pieces:

Theorem (3.1′)**.** *If* $\overline{\mathbb{B}}_1^3$ *and* $\overline{\mathbb{B}}_2^3$ *are disjoint unit balls, then* $\overline{\mathbb{B}}_1^3 \overset{10}{\simeq} \overline{\mathbb{B}}_1^3 \cup \overline{\mathbb{B}}_2^3$.

Proof: Let $\overline{\mathbb{B}}_1^3$ and $\overline{\mathbb{B}}_2^3$ have centres O_1 and O_2, and surfaces S_1^2 and S_2^2.

Let $F = \mathcal{F}\langle\alpha,\beta\rangle$ be a free group of rotations about O_1 (by Thm 2.3).

Let $P = \{p \in S_1^2 \mid \exists f \in F,\ \ fp = p\}$ = the set of points of S_1^2 on axes of rotations in F. Note that P is countable and no point of $P' = S_1^2 \setminus P$ is fixed by any rotation $f \in F$.

Define $x \sim y$ on P' by $\exists f \in F : x = fy$, an equivalence relation. Choose one point from each equivalence class (orbit) to form the set Q. Then each $x \in P'$ has a unique representation of the form fq, where $f \in F$ and $q \in Q$.

First, observe algebraically that $\widehat{P'} = F\widehat{Q}$ is partitioned by the sets

$$\{\widehat{Q},\ \ \omega(\alpha)\widehat{Q},\ \ \omega(\alpha^{-1})\widehat{Q},\ \ \omega(\beta)\widehat{Q},\ \ \omega(\beta^{-1})\widehat{Q}\}$$

Now, we have the following algebraically simple situation that is geometrically rather counterintuitive. The $\omega(f)\widehat{Q}$'s are "dense" physical objects that are now being rotated in ways that make some of them seemingly "disappear":

$$\begin{aligned}
\omega(\beta)\widehat{Q}\ \cup\ \omega(\beta^{-1})\widehat{Q}\ \cup\ \widehat{Q}\ &\underset{\alpha}{\overset{1}{\preceq}}\ \omega(\alpha)\widehat{Q}\\
\omega(\alpha)\widehat{Q}\ \cup\ \omega(\alpha^{-1})\widehat{Q}\ &\underset{\beta}{\overset{1}{\preceq}}\ \omega(\beta)\widehat{Q}, \text{ and the union gives}\\
\widehat{P'}\ =\ F\widehat{Q}\ &\overset{2}{\preceq}\ \omega(\alpha)\widehat{Q}\ \cup\ \omega(\beta)\widehat{Q}.
\end{aligned}$$

Hence, since P is countable, Corollary 2.5 gives

$$\begin{aligned}
\widehat{S_1^2} \overset{2}{\preceq} \widehat{P'}\ &\overset{2}{\preceq}\ \omega(\alpha)\widehat{Q}\ \cup\ \omega(\beta)\widehat{Q}, \quad \text{so by Thm 1.3,}\\
\widehat{S_1^2}\ &\overset{4}{\preceq}\ \omega(\alpha)\widehat{Q}\ \cup\ \omega(\beta)\widehat{Q}, \quad \text{and by symetry,}\\
\widehat{S_2^2}\ &\overset{4}{\preceq}\ \omega(\alpha^{-1})\widehat{Q}\ \cup\ \omega(\beta^{-1})\widehat{Q}.
\end{aligned}$$

Also, all the maps arranging $\widehat{S_1^2}$ into $\omega(\alpha)\widehat{Q} \cup \omega(\beta)\widehat{Q}$ are rotations about O_1 and thus leave O_1 fixed. Hence, we may adjoin O_1 to one of the pieces involved to obtain

$$\begin{aligned}
\overline{\mathbb{B}}_1^3 = \widehat{S_1^2} \cup \{O_1\} &\overset{4}{\preceq} \omega(\alpha)\widehat{Q} \ \cup \ \omega(\beta)\widehat{Q} \ \cup \ \{O_1\}, \quad \text{along with} \\
\overline{\mathbb{B}}_2^3 \setminus \{O_2\} = \widehat{S_2^2} &\overset{4}{\preceq} \omega(\alpha^{-1})\widehat{Q} \ \cup \ \omega(\beta^{-1})\widehat{Q}, \quad \text{and by translating } O_2 \text{ into } \widehat{Q}, \\
\{O_2\} &\overset{1}{\preceq} \widehat{Q}.
\end{aligned}$$

Taking the union of the arrangements above gives:

$$\begin{aligned}
\overline{\mathbb{B}}_1^3 \cup \overline{\mathbb{B}}_2^3 &\overset{9}{\preceq} F\widehat{Q} \cup \{O_1\} \subseteq \overline{\mathbb{B}}_1^3, \quad \text{hence} \\
\overline{\mathbb{B}}_1^3 \cup \overline{\mathbb{B}}_2^3 &\overset{9}{\preceq} \overline{\mathbb{B}}_1^3, \quad \text{and by the identity map,} \\
\overline{\mathbb{B}}_1^3 &\overset{1}{\preceq} \overline{\mathbb{B}}_1^3 \cup \overline{\mathbb{B}}_2^3, \quad \text{therefore, by Thm 1.3,} \\
\overline{\mathbb{B}}_1^3 \cup \overline{\mathbb{B}}_2^3 &\overset{10}{\simeq} \overline{\mathbb{B}}_1^3. \qquad \square
\end{aligned}$$

Although perhaps nothing can be surprising at this point, this result can be still further strengthened, giving a result that almost anyone would instinctively reject:

Theorem (3.2)**.** *(The Banach-Tarski Theorem, strong form) If T_1 and T_2 are bounded subsets of $\mathbb{R}^3$ with non-empty interior, then $T_1 \simeq T_2$ (using finitely many pieces).*

Proof: Let $B \subseteq T_1$ be a ball. Cover T_2 with balls $B_0, B_1, \ldots, B_n$ the same size as B (allowing overlaps). Let $B'_0, \ldots, B'$ be disjoint translates of $B_0, \ldots, B_n$, so clearly

$$\cup B_i \preceq \cup B'_i.$$

But $\cup B'_i \simeq B$ by induction from Theorem 3.1, hence

$$T_2 \subseteq \cup B_i \preceq \cup B'_i \simeq B \preceq T_1$$

and by symetry, $T_1 \preceq T_2$, hence $T_1 \simeq T_2$. $\square$

This would probably have been the most dramatic place to end our discussion of the theorem, however, a more indepth analysis is the aim of this paper. This theorem directly contradicts the usual expectation that rigid motion does not distort the volume of objects, and is in this sense a paradox. To "resolve" a paradox is to come to understand how ones initial expecations were unwarranted, or to realize that they have in fact been met. Here are some points worth noting for this purpose:

1) Some claim that the pieces involved in the proof have not been described by a "constructive" argument. The Axiom of Choice is used when choosing the points of the transversal Q, and some mathematicians reject the Axiom of Choice as being intuitively obvious. In fact, some use the BTT directly as an argument to refute the use of the Axiom of Choice.

However, it is worth noting that *arrangements in* $\mathbb{R}^2$ *actually preserve areas* (when the sets involved are measurable), and that the only known proofs of this statement employ the Axiom of Choice! Hence, although rejecting the Axiom of Choice would thwart unexpected results in $\mathbb{R}^3$, it would also result in the loss of the corresponding *expected* results in $\mathbb{R}^2$! I personally am not satisfied with this resolution, particularly since I advocate use of the Axiom of Choice.

2) An actual ball in the physical universe is comprised of a finite number of discrete particles and mostly empty space. The unit ball in Euclidean space, however, is completely dense and contains uncountably many points, a property which was necessary in the proof of the BTT. In this sense, the objects we have been arranging can be considered inappropriate as idealizations of physical balls. The maps in the proof could more appropriately be viewed to be arrangements of *empty space* rather than matter.

3) This paper proposes a new perspective: the isometries involved in the Theorem account only for the initial and final positions of the pieces, and do not describe the "trajectories" they follow from the initial ball into the final two balls. It is thus conceivable that the pieces are "interlocked" in such a way that they could never be separated (in the way that "Chinese linking rings" should be impossible to separate).

The remainder of this paper is therefore aimed at providing a context for the question of whether a Euclidean unit ball can be physically separated into a finite number of pieces and then reassembled to form two complete balls of the same size.

5 Movement

We now give an axiomatic treatment of movement in space. These axioms may be used to decide whether the pieces of the ball involved in proof of the BTT are actually physically separable. (An extensive justification of this definition can be found in my B.Sc. Honours thesis.)

Definition (5.4)**.** *(movement)*[3]

- A *movement* of $\mathbb{R}^n$ is a continuous map $\mu : [0,1] \to E(\mathbb{R}^n)$ such that $\mu(0) = I$. (all points are unmoved at time 0)
- The set of movements of $\mathbb{R}^n$ will be denoted by $\text{mov}(\mathbb{R}^n)$.
- The isometry $\mu(1)$ is the *terminal isometry* of μ.

As a consequence of this definition, it is relatively simple to show that movements preserve orientation:

Theorem (5.5)**.** *If μ is a movement, then* $\det(\mu(t))$ *is always* 1, *(i.e. movements preserve orientation).*

Proof: Since $\det : \mathfrak{M}_n(\mathbb{R})(\mathbb{R}) \to \mathbb{R}$ is continuous and $\mu : [0,1] \to E(\mathbb{R}^n)$ is continuous, it follows that $\det \circ \mu : [0,1] \to \mathbb{R}$ is continuous, and moreover, assumes only values of ± 1. Since $\det \circ \mu(0) = \det(I) = 1$, by continuity, $\det(\mu(t)) = 1$ for all $t \in [0,1]$. □

[3]There is a superficial resemblance between this and the definition of a one-parameter subgroup, but be warned that the property $\mu(a+b) = \mu(a) \circ \mu(b)$ is *not* true in general. A movement could start out spinning, then sliding, then stop, and then spinning and turning at the same time!

6 Reassemblies/the Banach-Tarski Problem.

Arrangements, like isometries, only specify initial and final positions of objects, and do not specify any intermediate positions, or the object "trajectories". Movements were defined to specify intermediate trajectories for objects undergoing isometries, and a concept of *reassemblies*[4] can be rigorously defined to serve the analogous purpose for arrangements.

The relationships between objects with respect to reassemblies can be developed fomrmally in a similar way to their relationships under arrangements relations (where, for example, composition of isometries is replaced by succession of movements, as definied in §5). However, an important difficulty is that set theoretic arguments like those used in proving the Banach-Schröder-Bernstein Theorem (Theorem 1.3, for arrangements) may produce pieces which are interlocked in some way which makes them physically inseparable (like jigsaw pieces in 2D). Thus, if A can be reassembled (taken apart without superimposing pieces) to form part of B, and B reassembled to form part of A, it is not clear whether A can be taken apart to form precisely B.for arrangements.

Therefore, we will bypass the less fruitful formal approach of defining reassemblies as ordered pairs (like arrangements) with composition and so on, since this paper is chiefly interested only in posing the following question:
Problem (6.3) *The Banach-Tarski Question (BTQ) in $\mathbb{R}^n$:*

> Is it possible to *reassemble* a single unit ball to form two unit balls? In other words, is there a partition $\mathcal{P} = \{P_1, \ldots, \mathcal{P}_k\}$ of $\mathbb{B}_0^n$, with movements $\mu_1, \ldots, \mu_k$ such that $\mu_i(t)[P_i]$ are always (for all t) disjoint and $\bigcup_i \mu_i(1)[\mathcal{P}_i] = \mathbb{B}_0^n \cup \mathbb{B}_1^n$ (disjoint unit balls)?

Since the original Banach-Tarski Theorem is known to be false in $\mathbb{R}^n$ for $n = 1, 2$, the answer to the BTQ in these cases is already "no." Also, the answer for $n > 4$ can be easily found, as follows. *To help visualize this process, think of 3-space as the plane and higher dimensions as vertical dimensions.*

1) Extend all isometries used in the proof of the BTT for $n = 3$ (see Thm 3.1) to isometries in $\mathbb{R}^n$ which fix coordinates in the 4th dimension and higher ("vertical coordinates").

2) Identify each point in each "hyperball" with the point in $\mathbb{R}^3$ vertically aligned with it (i.e., having the same first three coordinates), and likewise extend the pieces of the balls in 3-space to include the points vertically aligned with them.

3) Separate the pieces by (continuous) vertical translations (by various distances along the 4rth coordinate axis) into different regions of hyperspace.

[4]The prefix "re" is suggestive of the everyday notion that, in order to *re*assemble something, one must first be able to *dis*assemble it. In other words, its pieces not must be interlocked in some way that prevents them from being separated or "pulled apart."

4) Finally, apply the isometries of (1), and reunite the pieces by (continuously) reversing the vertical translations to form the two new balls.

Hence, only the $n = 3$ case of the BTQ is left unanswered. The remainder of this paper provides a partial answer to this question by showing that the partition pieces used to prove the BTT in §3 are in fact inseparable. Furthermore, the methods used are sufficiently general that they can be adapted to other BTT proofs, and at present, *all other BTT proofs explored by this author* (of which there are several) use pieces which are interlocked in a way that cannot be easily avoided.

7 Curve-density on spheres.

Our plan is to show that the sectoids (pieces) used in the BTT proof (which are arranged to form two copies of the unit ball from one) are somehow ″too dense″ to be pulled apart, thus avoiding the paradox that a ball can actually be ″taken apart″ to make two balls. However, the usual topological notion of a dense set will be too weak for this purpose: one can construct disjoint sectoids which are locally dense and cover S^2 but can be translated and rotated apart.

For example, take A_1,A_2 to be the subsets of the upper/lower hemispheres (resp.) with rational x coordinates, and let B_1,B_2 be their complements in their respective hemispheres. Then $\widehat{A_1}$ and $\widehat{B_1}$ are both locally dense but can be separated by vertically translating $\widehat{A_2}$ and $\widehat{B_2}$ the other two pieces away from them, and then rotating $\widehat{B_1}$ through 180° about the x-axis.

As a stronger alternative to topological density, we introduce a notion of ″curve-density.″

Definition (7.1). *Curve-Density*

- A *curve* will be defined as a non-constant continuous map from an interval $[a, b]$ into a manifold (for us, a sphere). We will often refer to a curve and its range (a subset of the manifold) interchangeably, and use the same symbol for each (the meaning should be clear from context).
- For sets $\mathcal{D}$, $\mathcal{E}$ in a manifold, $\mathcal{D}$ will be called *curve-dense on* $\mathcal{E}$ if every curve in $\mathcal{E}$ (that is, every curve *into* $\mathcal{E}$) necessarily intersects $\mathcal{D}$. For example, $\mathbb{R}^2 \setminus \mathbb{Q}^2$ is curve-dense (on $\mathbb{R}^2$).
- A set can be called near-curve-dense if it is curve dense when united with a countable set. For example, $\mathbb{R} \times \mathbb{Q}$ is near-curve-dense (on $\mathbb{R}^2$).

An intuitive example of a curve-dense set is a set that contains *so many curves* in *so many locations and orientations* that no other curve can possibly pass through such a tangled web of blockages. This is precisely the idea behind the following theorem, whose proof is somewhat messy, but naturally motivated. The result relates topological density in $SO(3)$ to curve-density on S^2 in a simple way that is easily applicable to the analysis of BTT proofs that use sectoids and paradoxical group partitions (as this paper does).

Theorem (7.2). *If Γ and Δ are curves on S^2, and a set of rotations $\mathcal{D}$ is dense in $SO(3)$, then Δ necessarily intersects the set* $\mathcal{D}\Gamma := \bigcup_{\delta \in \mathcal{D}} \delta[\Gamma]$. *In other words:*

$$\boxed{\text{If } \overline{\mathcal{D}} = SO(3), \text{ then } \mathcal{D}\Gamma \text{ is curve dense on } S^2.}$$

Proof: We will indicate the distance between two points A and B by $|A, B|$. Parametrize $\Delta : [a, b] \to S^2$ and $\Gamma : [p_1, q] \to S^2$. If either Γ or Δ is space filling (has nonempty two-dimensional interior) then the result follows directly from topological density, so assume both curves have empty interior. The intuitive idea behind the proof is as follows:

- Choose very close points D_1, D_2 on "opposite sides" of Δ, and the earliest point $\Gamma(p_2)$ on Γ such that $|\Gamma(p_1), \Gamma(p_2)| = |D_1, D_2|$.
- Find a rotation $\delta \in \mathcal{D}$ such that $\delta\Gamma(p_i) \approx D_i$, so that $\delta\Gamma$ is forced to cross over Δ when passing from $\delta\Gamma(p_1)$ to $\delta\Gamma(p_2)$.

The difficulty lies in restricting the situation so that $\delta\Gamma$ cannot possibly "go around" Δ in the second step. To do this, we will direct our attention to a segment of Γ which is long enough to reach from one side of Δ to the other, but too short to circumvent the endpoints of Δ. We now proceed with the details:

Choose $c \in (a, b)$, and choose a positive radius $r < |\Delta(c), \Delta(a)|, |\Delta(c), \Delta(b)|$. Let Ω be a circle of radius r centered at $\Delta(c)$.

Since $\Delta(a), \Delta(b)$ are outside Ω and $\Delta(c)$ is inside (at the centre), $\Delta(t)$ must be on Ω for some $t \in (a, c)$ and for some $t \in (c, b)$. Let $a_0 = \sup\{t \in (a, c) : \Delta(t) \in \Omega\}$ and $b_0 = \inf\{t \in (c, b) : \Delta(t) \in \Omega\}$, so by continuity, $\Delta(a_0), \Delta(b_0) \in \Omega$ and the curve $\Delta_0 := \Delta|_{[a_0,b_0]}$ is otherwise entirely inside Ω. Thus, Δ_0 divides the inside of Ω into two (at least) two open sets U_1 and U_2 which share $\Delta(c)$ as a common boundary point (this follows from the Jordan Curve Theorem[5]).

For $i = 1, 2$, choose $D_i \in U_i$ such that $|\Delta(c), D_i| < \frac{1}{2}|\Gamma(p_1), \Gamma(q)|, \frac{1}{3}r$, and set $d = |D_1, D_2| \leq |\Delta(c), D_1| + |\Delta(c), D_2|$. Then:

1) $|\Gamma(p_1), \Gamma(q)| > d > |\Gamma(p_1), \Gamma(p_1)| = 0$, so by continuity we may choose the parameter $p_2 = \inf\{t \in (p_1, q) : |\Gamma(p_1), \Gamma(t)| = d\}$, and we have

$$|\Gamma(p_1), \Gamma(p_2)| = d \;\text{ and }\; |\Gamma(p_1), \Gamma(t)| < d \;\text{ for } t \in [p_1, p_2)$$

2) $|\Delta(c), D_1| + d < r$, so we may choose $\varepsilon > 0$ such that the following hold:

$$|\Delta(c), D_1| + \varepsilon + d < r$$

$$\mathcal{B}_\varepsilon(D_i) \subseteq U_i,$$

where each $\mathcal{B}_\varepsilon(D_i)$ is an open ball of radius ε about D_i.

Now, choose a rotation $\alpha \in SO(3)$ so that $\alpha\Gamma(p_i) = D_i$ (this is possible since $|D_1, D_2| = d = |\Gamma(p_1), \Gamma(p_2)|$), and a sequence of rotations in $\mathcal{D}$, $\delta_n \to \alpha$, so that $\delta_n\Gamma(p_i) \to \alpha\Gamma(p_i) = D_i$. Next, choose $\delta = \delta_N$ for some sufficiently large N such that $\delta\Gamma(p_i) \in \mathcal{B}_\varepsilon(D_i) \subseteq U_i$, so $\delta\Gamma(t)$ intersects the boundary of U_1 for some $t \in [p_1, p_2]$. But we have

$$\begin{aligned} |\Delta(c), \delta\Gamma(t)| &\leq |\Delta(c), D| + |D, \delta\Gamma(p)| + |\delta\Gamma(p), \delta\Gamma(t)| \\ &= |\Delta(c), D| + |D, \delta\Gamma(p)| + |\Gamma(p), \Gamma(t)| \\ &\leq |\Delta(c), D| + \varepsilon + d \;<\; r, \end{aligned}$$

so $\delta\Gamma(t) \notin \Omega$ and we must have $\delta\Gamma(t) \in \Delta_0$, thus $\mathcal{D}\Gamma$ intersects Δ. □

[5]The case where Δ is nowhere injective can be remedied using an approximating differentiable curve

Having established this result, we are now motivated to search for useful sufficient conditions for a set of rotations to be dense in $SO(3)$, which will help us to construct and identify curve-dense sets. For this, we introduce some concepts and make use of two lemmas:

Notation/Terminology

- A *rational rotation* will be defined as a rotation whose angle is a rational multiple of 2*pi*, and other rotations will be called *irrational rotations*. Thus, a rotation is rational/irrational according as it is of finite/infinite order, according as it corresponds (canonically) to a rational/irrational element in the group $\mathbb{Q}/\mathbb{Z}$.
- The set of rotations in $SO(3)$ about an axis v will be denoted by $SO(3)_v$. Since a point $p \in R^3$ determines a unique axis through the origin, $SO(3)_p$ will denote the set of rotations about that axis, i.e., the elements of $SO(3)$ which leave p fixed.

Lemma (7.3). *If α is an irrational rotation about an axis v, then the topological closure of the natural powers of alpha contains all rotations about v. That is,*

$$\overline{\alpha^{\mathbb{N}}} = SO(3)_v$$

Proof: This is equivalent to the property that, for irrational x, the set $\mathbb{N}x$ in the group $\mathbb{Q}/\mathbb{Z}$ has no element of minimal absolute value, which can be shown using a Euclidean Algorithm based argument. □

Lemma (7.4). *The topological closure of a semigroup in $SO(3)$ is a Lie subgroup.*

Proof: Let the semigroup be $\mathbb{S}$. We know that $\overline{\mathbb{S}}$ is operation-closed, by continuity, and so it remains only to show that $\overline{\mathbb{S}}$ is closed under inversion. If $\alpha \in \overline{\mathbb{S}}$ is a rational rotation, then α^{-1} is a finite power of α which is in $\overline{\mathbb{S}}$. If α is irrational about an axis v, then $\overline{\mathbb{S}} \supseteq \overline{\alpha^{\mathbb{N}}} = SO(3)_v$ which contains α^{-1}. Thus $\overline{\mathbb{S}}$ is a subgroup which is topologically closed, and so is a Lie subgroup. □

Theorem (7.5). *If a semigroup $\mathbb{S} \subseteq SO(3)$ contains two non-coaxial irrational rotations, then it is dense in $SO(3)$, i.e., $\overline{\mathbb{S}} = SO(3)$.*

Proof: Let the rotatations be α and β, and let their axes be u and v, respectively. We know by Lemma 7.3 and Lemma 7.4 that $\overline{\mathbb{S}}$ is a Lie subgroup of containing $SO(3)_u$ and $SO(3)_v$, and so its Lie algebra contains linearly independent vectors U and V derrived from these one-parameter subgroups. But their commutator $[U, V]$ is a third linearly independent vector[6], so the Lie algebra of $\overline{\mathbb{S}}$ is of the same vector dimension as $so(3)$ (the Lie algebra of $SO(3)$), and must

[6] To see this, observe that $so(3)$, the 3×3 skew symetric real matrices under the Lie bracket, is isomorphic to $\mathbb{R}^3$ under the cross product, and the cross product of two linearly independent vectors is a third linearly independent vector.

therefore *be* $so(3)$. Since $SO(3)$ is completely determined by its Lie algebra, it follows that $\overline{\mathcal{S}} = SO(3)$. $\square$

Now it is time to apply these results to sets involved in this paper's proof of the BTT. See Theorem 3.1 to review the definitions of the sets.

Theorem (7.6)**.** *If A is one of $\omega(\alpha)Q$, $\omega(\alpha^{-1})Q$, $\omega(\beta)Q$, $\omega(\beta^{-1})Q$, then $A \cup P$ is curve-dense on S^2 (with terms defined as in the proof of Theorem 3.1)*

Proof: Suppose WLOG that $A = \omega(\alpha)Q$ and that $\omega(\alpha)Q \cup P$ is not curve-dense, so $\Gamma \cap (\omega(\alpha)Q \cup P) = \emptyset$ for some curve Γ, i.e.

$$\Gamma \subseteq S^2 \setminus (\omega(\alpha)Q \cup P) \subseteq (F \setminus \omega(\alpha))Q \cup P$$

Since $\alpha\beta\alpha^{-1}(F \setminus \omega(\alpha)) \subseteq \omega(\alpha)$ and $\alpha\beta\alpha^{-1}P = P$, we have $\alpha\beta\alpha^{-1}\Gamma \subseteq \omega(\alpha)Q \cup P$. Now let $\mathcal{C} = \{\gamma \in F : \gamma\omega(\alpha) \subseteq \omega(\alpha)\}$. Since $\mathcal{C}$ is a semigroup containing non-coaxial irrational rotations $\alpha\beta$ and $\alpha^2\beta$, it follows that $\mathcal{C}$ is dense in $SO(3)$, and thus $\mathcal{C}\alpha\beta\alpha^{-1}\Gamma \subseteq \omega(\alpha)Q \cup P$ are both curve-dense on S^2. $\square$

The results leading up to Theorem 7.6 (excluding Theorem 7.1) were mostly simple enough to be combined into a single proof. However, they were given separately and in sufficient generality as to be useful in analyzing other BTT proofs, which employ rotational arrangements of the sphere that can be extended to sectoids.

8 Inseparability of sectoids.

To define what it means for certain pieces of a ball to be "impossible to take apart" or "inseparable," a general definition of reassembly is not needed.

Definition (8.1). *(Separability of Sectoids)*

- A collection of disjoint concentric[7] sectoids $\widehat{A_1}, \ldots, \widehat{A_n}$ with common centre O will be called *separable* if there exist *separating movements* $\mu_1, ..., \mu_n$ (over a common time interval $[0,1]$) such that the points $\mu_i(1)[O]$ are distinct and the sets $\mu_i(t)[\widehat{A_i}]$ are disjoint $\forall t$ (in $[0,1]$). (This relation is clearly symetric among the sectoids, and does not entail the separability of smaller subcollections of the n sectoids.)
- For existence purposes, we may assume that $\mu_1 = I$ ($\widehat{A_1}$ does not move) by replacing $\mu_i(t)$ by $\mu_i(t)\mu_1(t)^{-1}$.
- A collection of disjoint concentric sectoids will be called *inseparable* if they are not separable.

As an example in two dimensions, the pieces of an assembled jigsaw puzzle are inseparable (because they are "interlocked"). The ship and bottle of a ship-in-a-bottle are not necessarily inseparable in the above sense because it may be possible to "jiggle" the ship inside the bottle (one may wish to define a notion of "partial separability" for this scenario).

It is clear that if the answer to the BTQ is 'yes,' and involves only sectoid pieces of the balls (ignoring the centres), then the collection of sectoids must be separable.

Convention: *The unit sphere, (open) ball, and sectoids centered at a point c will be denoted by* $\mathbb{B}^3_c$, S^2_c *, etc., with no subscript if c is the origin.*

If $\widehat{A}, \widehat{B}$ are unit sectoids centered at the origin and μ separates $\widehat{B}$ from $\widehat{A}$, the centre $c = \mu(t)(0)$ of $\mu(t)[\widehat{B}]$ is *initially* inside the unit ball $\mathbb{B}^3$. Therefore, *we will henceforth always restrict ourselves to 'small scale' movements* μ *such that, for all* t, $c = \mu(t)(0) \in \mathbb{B}^3$.

We now investigate sufficient conditions to show that two sectoids $\widehat{A}, \widehat{B_c}$ with *nearby centres* $0, c \in \mathbb{B}^3$ are non disjoint. There are easy criteria of this sort which examine subsets of the spheres S^2, S^2_c, since an intersection point p of $\widehat{A}$ and $\widehat{B_c}$ can be radially projected to a point $p_1 \in A \subseteq S^2$ or to a point $p_2 \in B_c \subseteq S^2_c$.

There are many ways of choosing onto which sphere and along which radii to make these projections, but one method derrived from this line of thinking results in a homeomorphism $S^2_c \to S^2$, making it particularly amenable with the curve density results of §7:

Definition (8.2). *(the Radial Homeomorphism for proximal spheres)*

[7] Recall that a sectoid does not contain its centre as a point.

Given a point $c \in \mathbb{B}^3$, there is a homeomorphism $h_c : S_c^2 \to S^2$ given by:

$$x \quad \mapsto \quad \begin{cases} \overline{0x} \cap S^2, & x \in \mathbb{B}^3 \\ \overline{cx} \cap S^2, & x \notin \mathbb{B}^3 \end{cases}$$

That is, depending on where x is, it maps to the intersection of a certain radial segment with S^2 (taken as a point rather than a singleton). Notice that if $x \in S^2 \cap S_c^2$, then $h_c(x) = x$. In particular, if $c = 0$, then h_c is the identity.

Now, observe (geometrically) that

$$\text{for } x \in S_c \cap \mathbb{B}^3 \text{, we have} \qquad x \in \widehat{A} \cap B_c \iff h_c(x) \in A, x \in B_c, \text{ and}$$

$$\text{for } x \in S_c \setminus \mathbb{B}^3 \text{, we have} \quad h_c(x) \in A \cap \widehat{B_c} \iff h_c(x) \in A, x \in B_c, \text{ hence}$$

$$h_c(x) \in A,\ x \in B_c \ \Rightarrow\ x \in \widehat{A} \cap B_c \text{ or } h_c(x) \in A \cap \widehat{B_c} \ \Rightarrow\ \widehat{A} \cap \widehat{B_c} \neq \emptyset, \text{ therefore}$$

$$\widehat{A} \cap \widehat{B_c} = \emptyset \ \Rightarrow\ A \cap h_c[B_c] = \emptyset$$

This relationship was precisely the motivation for the definition of h_c. It allows us to show that two sectoids intersect by showing that two subsets of the sphere intersect. If $\widehat{A}$ and $\widehat{B}$ are disjoint unit sectoids (centered at the origin), and μ is a movement such and $\widehat{A} \cap \mu(t)[\widehat{B}] = \emptyset$, then by the above discussion,

$$A \cap h_{\mu(t)(0)}[\mu(t)[B]] = \emptyset$$

which, by letting $h^\mu(t) = h_{\mu(t)(0)} \circ \mu(t)$, can be written as

$$A \cap h^\mu(t)[B] = \emptyset$$

Notice that if $\mu(t)$ fixes the origin and thus takes $S^2 \to S^2$, then $h^\mu(t) = \mu(t)$ on S^2, so in particular, $h^\mu(0)$ is always the identity. For nominal completeness, we call h^μ the **Radial Tracer** for μ, and summarize the above in the following

Theorem (8.3). *(Radial Tracing Criterion for proximal sectoid movement)*
For unit sectoids given by $A, B \subseteq S^2$ and a movement μ,

$$\boxed{\forall t, \quad \widehat{A} \cap \mu t[\widehat{B}] = \emptyset \Rightarrow A \cap h^\mu(t)[B] = \emptyset}$$

The Radial Tracer h^μ has been so named because, as B moves "through A", we trace out a curve $h^\mu(t)(b)$ on S^2 for each $b \in B$, and these curves must be disjoint from A if $\widehat{B}$ remains disjoint from $\widehat{A}$. The Radial Tracing Criterion will be used to show the inseparability of the sectoid pieces used in our proof of the BTT (Theorem 3.1), and applies to other BTT proofs which use sectoids.

Another characteristic which is common among sectoidal proofs of the BTT is that the pieces used to partition the sphere *must be non-measurable*, and tend to exhibit other pathological behaviour, such as containing disjoint isometric copies of themselves and/or each other. Now, the *inner-measure* of a set, A, is defined as

$$\underline{m}(A) := sup\{m(X) | X \subseteq A, X \text{ measurable}\}$$

If A and B are disjoint, then $\underline{m}(A \cup B) \geq \underline{m}(A) + \underline{m}(B)$ (inner-measure is superadditive over disjoint unions), since if X, Y are measurable subsets of A, B (respectively), then $X \cup Y$ is a measurable subset of $A \cup B$ with measure $m(X) + m(Y)$.

The measure theoretic 'weirdness' of the pieces used in our BTT proof, and other proofs, tends to prevent them from containing measurable subsets. The argument for our proof is given explicitly below:

Lemma (8.4)**.** *(Inner-Measure Lemma for paradoxical the pieces used in Thm. 3.1)*
If A is a union of at most three of the sets $\omega(\alpha)Q$, $\omega(\alpha^{-1})Q$, $\omega(\beta)Q$, $\omega(\beta^{-1})Q$, then A is of inner-measure zero (as a subset of S^2).

(This property will then also be true of $A \cup P$ since P is of measure zero.)

Proof: In either of the above cases, (check that) A contains two disjoint isometric copies of itself, (e.g., $\omega(\alpha)Q$ contains $\alpha\beta\omega(\alpha)Q$ and $\alpha^2\beta\omega(\alpha)Q$), so the inner-measure of A is at least twice the inner-measure of A. Since A cannot be of infinite inner-measure (being a subset of S^2), it must be of inner-measure zero. □

At long last, we are ready to prove the inseparability of our sectoids:

Theorem. *Suppose A and B are any two of $\omega(\alpha)Q$, $\omega(\alpha^{-1})Q$, $\omega(\beta)Q$, $\omega(\beta^{-1})Q$. Then, $\widehat{A}$ and $\widehat{B}$ are inseparable.*

Proof: Suppose there exists a movement μ separating B from A. Assume WLOG, that $A = \omega(\alpha)Q$, $B = \omega(\beta)Q$. Also restrict the t-domain of μ so that $\mu(t)(0) \in \mathbb{B}^3$. Then, by the Radial Tracing Criterion, it is necessary that $A \cap h^{\mu}(t)[B] = \emptyset$. In particular, we have $A \cap h^{\mu}(t)\beta Q = \emptyset$. However, for any particular q, since the curve $h^{\mu}(t)\beta q$ (traced out as t varies) is non-constant (with at most two exceptions[8]), it must intersect $A \cup P$ (by curve-density). Since the curve does not intersect A, it must intersect P.

The idea now is that there are "too many" curves $h^{\mu}(t)\beta q$ to all pass through P, which we can show by a measure argument. Notice that, by the Inner-Measure Lemma, any measurable subset of $S^2 \setminus A$ must be of measure zero, so in particular, the curves $h^{\mu}(t)\beta q$ are of measure zero (being measurable due to compactness). On the other hand, the set Q must *not* be of measure zero, since if it were, then $S^2 = FQ$ (a countable union of isometric copies of Q) would also of measure zero, a contradiction.

For each q, we have shown that $h^{\mu}(t)\beta q$ intersects P, i.e., for some t value, $h^{\mu}(t)\beta q \in P$, thus $q \in H(t)P$ by setting $H(t) = \beta^{-1}h^{\mu}(t)^{-1}$. Although a different t value may be necessary for each q, it is always contained in the interval of definition of h^{μ}, say I. Thus

$$Q \subseteq H(I)P$$

Now, the righthand side is a countable union of curves $H(I)p$ which are of measure zero, so Q is of measure zero, a contradiction. Hence, for any non

[8]It can be shown that at most two points on S^2 are fixed under the homeomorphisms h^{μ}, which can be ignored for simplicity because this proof is a based on a measure argument.

constant movement, μ, $A \cap h^{\mu}(t)[B] \neq \emptyset$ for some t, and thus $\widehat{A} \cap \mu(t)[\widehat{B}] \neq \emptyset$. In other words, $\widehat{A}$ and $\widehat{B}$ are inseparable. □

The main concepts involved in showing our sectoids to be inseparable were:

- the sufficient conditions for density in $SO(3)$ and curve-density on S^2;
- the near-curve-density of the sectoid surfaces;
- the Inner-Measure Lemma for the paradoxical pieces; and
- the Radial Tracing Criterion for proximal sectoid movement, which rests heavily on the fact that Radial Tracer, h^{μ}, is homeomorphism-valued, allowing us to take the inverse images of surface curves in an organized way to countably cover the transversal set Q.

The arguments will not be given here, but these ideas are sufficient to prove the inseparability of the pieces used in every BTT proof known to this author. These include adaptations from existing proofs which use a larger number of pieces obtained by restricting the old pieces to measurable subsets of the sphere, e.g. the north and south hemispheres.

These results suggest that the answer to the Banach-Tarksi Question is "no," which poses the following

Conjecture: *There is no reassembly of the unit ball to form two unit balls. Equivalently, any arrangement of the unit ball to form two unit balls will make use of interlocking pieces.* **That is, it is impossible to "take apart" a unit ball to make two unit balls.**

9 Concluding Remarks / References

In addressing the Banach-Tarski Question, the techniques of this paper are only known to show that *particular* constructions fail, and have not been applied to rule out the possibility that some construction may work. It seems plausible that perhaps all partitions of the unit ball into *sectoids* might fail to give a positive answer to the BTQ, although there is still a great deal of freedom in choosing the transversal set, Q, under the action of some ”paradoxical” subgroup of $SO(3)$ such as a free group.

If the approach of dividing the ball into sectoids is indeed bound to fail, the techniques of this paper might be useful in proving it. Unfortunately, though, the complexity of a sectoid compared with that of an arbitrary subset of $\mathbb{B}^3$ is infinitesimal. It is unclear how the results for sectoids would extend to general subsets in any useful way.

The general BTQ in $\mathbb{R}^3$ is therefore unanswered. Still, the common interpretation of the Banach-Tarski Theorem as ”taking apart” a unit ball (in $\mathbb{R}^3$) to make two unit balls has been shown to be inappropriate and unjustifiable using existing BTT proofs. In this sense, the Banach-Tarski Paradox has indeed been resolved, or at least significantly demystified.

Bibliography

[1] S. Wagon , *The Banach-Tarski Paradox* , Cambridge ; New York : Cambridge University Press, 1985.

[2] A.A. Sagle, R.E. Walde , Introduction to Lie groups and Lie algebras , New York, Academic Press, 1973.