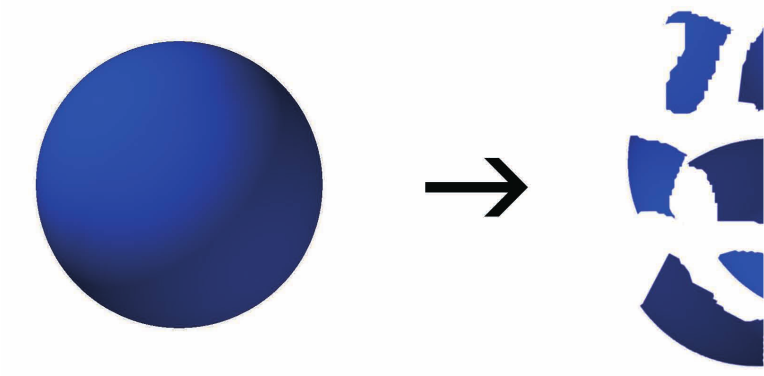
The Banach-Tarski

The Polish mathematicians Stefan Banach and Alfred Tarski published in 1924 [1] a proof of the following rather surprising fact:
A solid sphere can be cut into a finite number of pieces and reassembled to form two identical copies of itself.



Figuur 1: Illustration of the Banach-Tarski paradox.

DOOR GABRIEL CHÊNEVERT

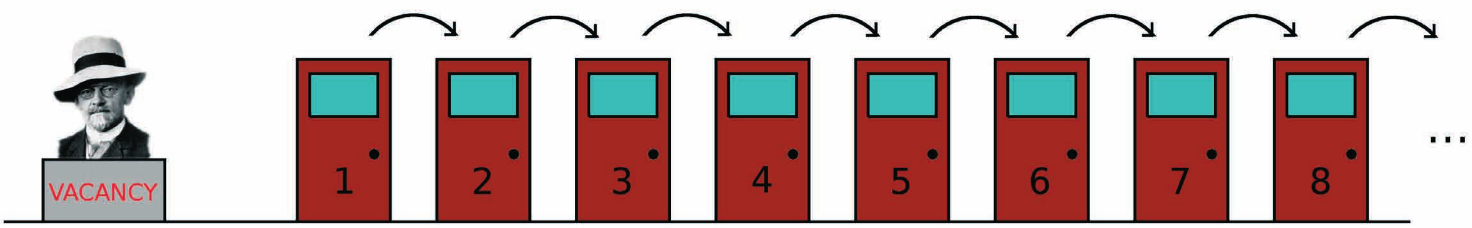
This result is at first glance quite at odds with geometrical intuition. Indeed, it seemingly allows one to create a ball out of thin air, using only another ball as a catalyst! But this is certainly not the only mathematical result that challenges common sense; let us first describe a simpler version of a similar phenomenon.

The Hilbert Hotel

Imagine a hotel consisting of a single (very) long corridor comprising infinitely many rooms, labeled by the natural integers. This hotel, named after David Hilbert (1862-1943), has the pleasant property of being able to accommodate new guests even when it is full! Indeed, suppose that all the rooms are occupied and that a new guest shows up at the front desk. The manager could ask all current occupants to simultaneously move to the next room: if the person in the first room moves to the second one, the person in the second moves to the third, and in general the occupant of room n moves to room $n + 1$, then the first room becomes available for the new guest without requiring any of the previous guests to leave the hotel (see figure 2). This feat is of course impossible to accomplish in a hotel with only finitely many rooms (such as those usually encountered in most contemporary cities), illustrating of the fact that an infinite set, unlike a finite one, can very well be in bijection with a proper subset of itself.

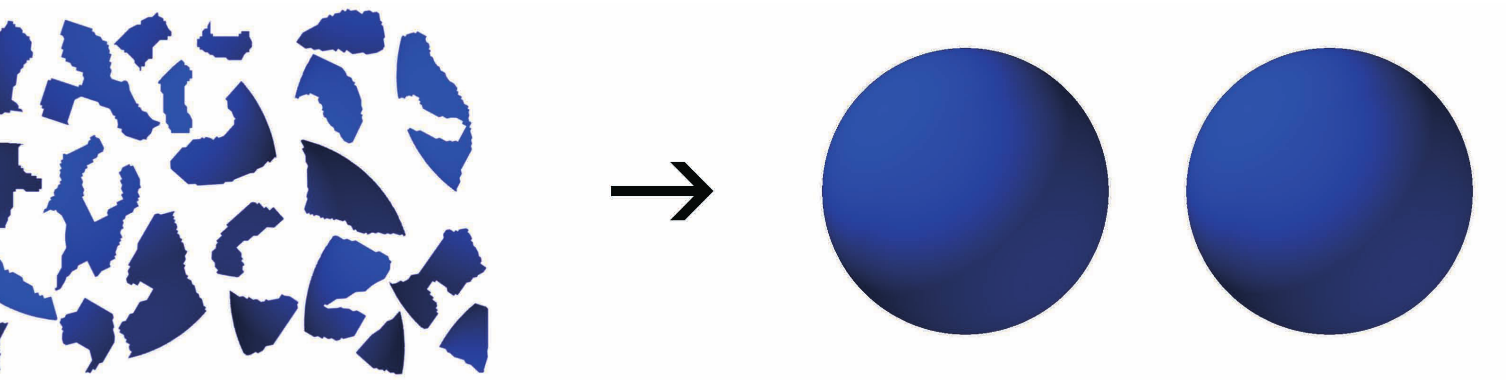
It is even possible to allow infinitely many new guests in the hotel by asking the occupant of room n to move to room $2n$, making all the odd-numbered rooms available. This means that the Hilbert Hotel can be "doubled": suppose that the two children of the manager each inherit half of the hotel at the manager's death; if the first one takes all the odd-numbered rooms and the second one all the even-numbered ones, they can both run their own Hilbert Hotel. Another illustration of this idea in popular culture is the joke of the fool, having been granted three wishes, who asks for "two other bottomless bottles of beer"!

Here is a geometric version of the same idea. Let C be a circle in the plane and suppose we want to fit one more point P in it. Since there are already infinitely many points on C, we only need to rearrange them in order to make room for one more.



Figuur 2: Plenty of room at the Hilbert Hotel.

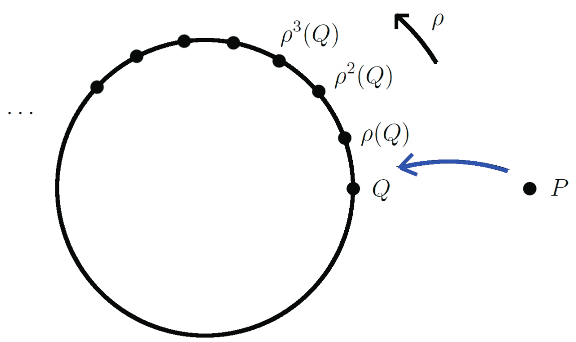
Paradox



If ρ denotes the rotation of $\sqrt{2}$ degrees around the center of C, then the successive images

$$Q,\ \rho(Q),\ \rho^2(Q),\ \rho^3(Q),\ \ldots$$

of any point $Q \in C$ under the iterates of ρ are all distinct. These points behave just like the rooms in the Hilbert Hotel: applying ρ make them "move to the next spot", leaving the location previously occupied by the point Q available for our extra point P. The only thing really needed in order for this trick to work is
That ρ have infinite order; hence any choice of an irrational rotation angle (in degrees) would work. Note that the set of iterates of Q is geometrically pretty wild: it is relatively small (i.e., only countably infinite) but nonetheless dense in C – just like the rational numbers on the real line.



Figuur 3: Making room for one more point on a circle

Equidecomposability

The statement made by Banach and Tarski actually concerns the group $E^+(3)$ of rigid motions (also called direct isometries) of 3-space. Recall that these are the bijections $\sigma : \mathbb{R}^3 \to \mathbb{R}^3$ that preserve distance and orientation; any such rigid motion can be obtained as a rotation followed by a translation (or vice versa). In fact, the rigid motions that leave the origin fixed are precisely the rotations around an axis going through the origin; they form a subgroup, denoted $SO(3)$, of $E^+(3)$. Two subsets of $\mathbb{R}^3$ are called congruent if one can be obtained by applying a suitable rigid motion to the other; for example, all cubes with the same volume are congruent, but none of them is congruent to any sphere.
We can now make precise the concept of "cutting and rearranging" alluded to in the introduction. We say that two sets A and B in $\mathbb{R}^3$ are equidecomposable if they can be written as finite disjoint unions of pair wise congruent subsets. In symbols, $A \sim B$ means that

$$A = \bigcup_{i=1}^{n} A_i, \quad B = \bigcup_{i=1}^{n} B_i, \quad \text{with } A_i \cap A_j = B_i \cap B_j = \varnothing \text{ whenever } i \neq j,$$

and for every $1 \le i \le n$ there is a rigid motion $\sigma_i \in E^+(3)$ such that $B_i = \sigma_i(A_i)$. This defines an equivalence relation weaker than congruence, in the sense that more pairs of subsets are now considered to be "the same" since we are allowed to apply a different rigid motion to each piece. For example, the Hilbert Hotel trick above shows that a circle in the plane is equidecomposable with a circle together with a point: if D stands for the set of iterates of Q, apply ρ to D in order to make room for P, then put P in place by applying to it the translation that sends it to Q. On the complement $C \setminus D$ of D we do nothing; this can be thought of as applying the trivial transformation (denoted 1 in the diagram below).

$$\begin{array}{ccccccc} C \cup \{P\} & = & C \setminus D & \cup & D & \cup & \{P\} \\ \wr & & \downarrow 1 & & \downarrow \rho & & \downarrow \sigma \\ C & = & C \setminus D & \cup & D \setminus \{Q\} & \cup & \{Q\} \end{array}$$

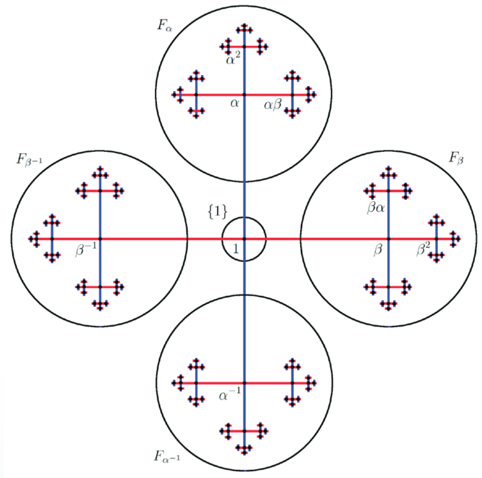
A set is called paradoxical if it equidecomposable with two disjoint copies of itself. Using this terminology, the theorem of Banach and Tarski can thus be stated as: a solid sphere is a paradoxical subset of $\mathbb{R}^3$.

Note that this precise formulation removes some of the aura of mystery around the "paradox". Certain variants of it are notably much easier to accept, for example if in addition to rigid motions one accepts other kinds of continuous deformations in the group of admissible transformations (we can easily imagine cutting a ball in two and inflating each piece to a ball of the original volume). We also have complete freedom to choose the pieces of the decompositions as we want. The bottom line here is that we should not be too troubled by certain subsets of R^3 not behaving according to our geometric intuition, because an arbitrary subset of R^3 does not usually look like something we might reasonably want to call a geometric object. In particular, the apparent paradox disappears once we realize there is a priori no reason why the usual notion of volume should apply to all subsets of $\mathbb{R}^3$ (and this theorem actually shows quite clearly that it does not).

Sketch of the proof

The crucial point that allows to construct a paradoxical decomposition of the ball is the fact that SO(3) contains two rotations α and β that are completely independent. For example, one could take α and β to be two rotations of angle $\arccos(\frac{1}{3})$ around two orthogonal axes[2]. In fact, just like in the 2D case where most rotations have infinite order, most pairs of rotations in SO(3) are independent, so just picking α and β randomly should do.

"Independent" here means that all the rotations obtained by performing various sequences of α, β and their inverses are distinct (if we convene in such a sequence never to undo a rotation immediately after performing it). In such a case we say that the subgroup F of SO(3) consisting of all these rotations is freely generated by α and β (see figure 4).



Figuur 4: A picture of the free group F generated by α and β. The vertices represent the elements of F, expressed as sequences made out of the symbols α, α^{-1}, β and β^{-1}. Going up a blue edge up corresponds to adding an α to the right of a sequence, and going down to removing it (or equivalently, adding an α^{-1}). Similarly, following a red edge right or left has the effect of adding a generator β or β^{-1}, respectively.

The group F exhibits a certain self-similarity that we can describe as follows. Write F as a disjoint union of five subsets

$$F = \{1\} \cup F_\alpha \cup F_{\alpha^{-1}} \cup F_\beta \cup F_{\beta^{-1}},$$

where the subscript denotes the leftmost generator appearing in an element of F (for example $\alpha^{-3}\beta^2\alpha \in F_\alpha{}^{-1}$, while $\beta^4\alpha^{-1} \in F_\beta$). If a sequence starts with α^{-1}, then according to our convention the second generator can be anything except α; adding an α on the left of such a sequence removes the α^{-1} and the second generator become the first in the process. In symbols,

$$\alpha F_{\alpha^{-1}} = \{1\} \cup F_{\alpha^{-1}} \cup F_\beta \cup F_{\beta^{-1}} = F \setminus F_\alpha,$$

and the same holds with α replaced by β. We thus obtain an equidecomposition

$$\begin{array}{ccccccccc} F \setminus \{1\} & = & F_\alpha & \cup & F_{\alpha^{-1}} & \cup & F_\beta & \cup & F_{\beta^{-1}} \\ \wr & & \downarrow 1 & & \downarrow \alpha & & \downarrow 1 & & \downarrow \beta \\ F \cup F & = & F_\alpha & \cup & F \setminus F_\alpha & \cup & F_\beta & \cup & F \setminus F_\beta \end{array}$$

of $F \setminus \{1\}$ with 2 copies of F. Using the Hilbert Hotel trick, one easily sees that $F \setminus \{1\} \sim F$, so we conclude that $F \sim F \cup F$, i.e., that F, as a group, is paradoxical.

Let now S be a sphere centered at the origin in R^3 and D the set of points in S that are fixed by some nontrivial rotation in F. Two points in $S \setminus D$ can be considered equivalent if we


Figuur 5: Stefan Banach (1892–1945) and Alfred Tarski (1901–1983)

can obtain one from the other by applying an element of F; this element is then unique since we made sure to take out the fixed points D. This allows to decompose $S \setminus D$ in infinitely many disjoint orbits for the action of F. The choice of a base point in an orbit identifies it with F; doing this for all orbits simultaneously, we obtain a paradoxical decomposition of $S \setminus D$ induced by that of F.
From this, it is relatively easy to deduce that S itself is paradoxical (we only took out a countable set of points after all). We may then get a paradoxical decomposition of a solid sphere B by considering it as the union of infinitely many concentric copies of S (like the layers of an onion, the origin has to be treated separately with one last application of the Hilbert Hotel trick). We thus obtain a proof of the fact that B is paradoxical.

Conclusion

Here we only sketched some the ideas behind the Banach-Tarski theorem, following loosely the modern account given in [2]. There has been over the years various refinements and variants of it – we now know that the smallest possible number of pieces in a paradoxical decomposition of the ball is five, and that the pieces can even be moved continuously in space without running into each other! Much attention has also been devoted to the crucial role played above by the axiom of choice which was used to simultaneously pick a base point in every orbit; the interested reader may wish to consult [3] for a survey of these developments.

Referenties

[1] S. Banach and A. Tarski, Sur la d´ecomposition des ensembles de points en parties respectivement congruentes, Fundamenta Mathematicae 6, pp. 244–277 (1924).
[2] S. Wagon, The Banach-Tarski paradox, Cambridge University Press (1985).
[3] The Banach-Tarski paradox, Wikipedia, http://en.wikipedia.org/wiki/Banach-Tarski paradox.

Mathematisch Instituut, Universiteit Leiden, Postbus 9512, 2300 RA Leiden, Nederland

Gabriel Chênevert
Gabriel Chênevert is currently a DIAMANT postdoc at the Mathematisch Instituut in Leiden. He was born in Montreal, Quebec and he received his Masters degree in mathematiscs in 2003 at the Université de Montréal. Afterwards he finished his Ph.D. in 2008 at the McGill University.

He works in the field of the arithmetic algebraic geometry and his main objects of study are Riemann zeta function of certain arithmetic varieties. In other words, he investigates the interplay of number theory and geometry. This article is developed out of an expository talk given at the Wiskunde Bachelorseminarium.