HOW TO TRISECT AN ANGLE
(Using P-Geometry)
(DRAFT: Liable to change)

Aaron Sloman
School of Computer Science, University of Birmingham
(Philosopher in a Computer Science department)

NOTE Added 30 Jan 2020
Remarks on angle-trisection without the neusis construction can be found in Freksa et al. (2019)
NOTE Added 1 Mar 2015
The discussion of alternative geometries here contrasts with the discussion of the nature of descriptive metaphysics in
"Meta-Descriptive Metaphysics: Extending P.F. Strawson's 'Descriptive Metaphysics'"
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-descriptive-metaphysics.html

This document makes connections with the discussion of perception of affordances of various kinds, generalising Gibson's ideas, in
http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#gibson
Talk 93: What's vision for, and how does it work? From Marr (and earlier) to Gibson and Beyond
Some of the ideas are related to perception of impossible objects.
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/impossible.html

---

JUMP TO CONTENTS

Installed: 26 Feb 2015
Last updated:
A very nice geogebra applet demonstrates the method described below:
http://www.cut-the-knot.org/pythagoras/archi.shtml.
Feb 2017: Added note about my 1962 DPhil thesis
25 Apr 2016: Fixed typo: ODB had been mistyped as ODE (Thanks to Michael Fourman)
29 Oct 2015: Added reference to discussion of perception of impossible objects.
4 Oct 2015: Added reference to article by O'Connor and Robertson.
25 Mar 2015: added (low quality) 'movie' gif showing arrow rotating.
2 Mar 2015 Formatting problem fixed.
1 Mar 2015 Added draft Table of Contents.
1 Mar 2015 Added appendix on the "Neusis construction"
28 Feb 2015 (Broke the trisection explanation into simpler steps);

This paper is available in two formats: html and pdf:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/trisect.html
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/trisect.pdf

A partial index of discussion notes is in
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/AREADME.html

This is part of the Turing-inspired Meta-Morphogenesis project, summarised here:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-morphogenesis.html

It also relates to an attempt to produce new Biological Evolutionary foundations for mathematics (BEFM) based on the roles of various kinds of construction kits in biological evolution. Natural selection depends on mathematical properties of construction kits, including the "Fundamental construction kit" provided by Physics+Chemistry, from which Evolution repeatedly derives and uses new construction kits. It thereby (blindly) discovers and proves mathematical theorems about what is possible, theorems that it puts to good use in creating instances of those possibilities in the form of new biological materials, mechanisms, functions, types of information, mechanisms for processing and using information, new species, and new products of interactions between species and their environments).
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/construction-kits.html
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/compositionality.html
There are also pdf versions of all the above documents.

CONTENTS

Preamble:
# What goes on when toddlers make topological discoveries

http://www.cs.bham.ac.uk/research/projects/cogaff/misc/toddler-theorems.html#holes

Background: The Triangle Sum theorem and P-geometry

Many years ago, a young mathematics school teacher, Mary Pardoe, visited me at Sussex University, where she had previously studied mathematics and philosophy. She told me about an unusual proof of the triangle sum theorem that she found her pupils could understand and remember more easily than the standard proof using parallel lines.

Her proof, using a rotated arrow, is compared with "standard" proofs here:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/triangle-sum.html

The sequence of pictures below (from left to right) summarises her proof.
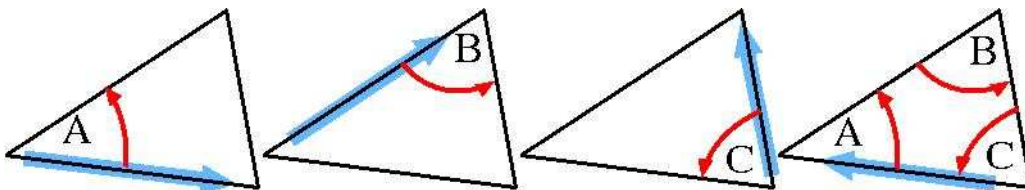
*It may be best to think of the proof not as a static diagram but as a **process**, with stages represented from left to right in the figure. (See "movie version" below). In the first stage, the pale blue arrow starts on the bottom side of the triangle, pointing to the right. It is then rotated through each of the internal angles A, B, C, always rotated in the same direction (counter-clockwise in this case), so that it lies on each of the sides in succession, until it is finally rotated through the third angle, C, after which it lies on the original side of the triangle, but obviously pointing in the opposite direction. Some people may prefer to rotate something like a pencil rather than imagining a rotation depicted by snapshots. The triangle can have any shape. It should be clear that rotation of the arrow about successive internal angles of the triangle will always bring it back to the original side, after half a rotation, no matter what the lengths of the sides. (Why?)*

MOVIE VERSION
(Works only in some web browsers)
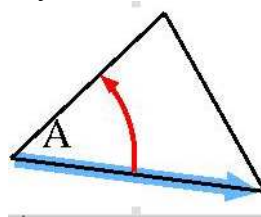
Unlike "standard" proofs of the theorem, this nowhere makes use of Euclid's parallel postulate, since no two lines are assumed to be parallel. It does assume that if two lines intersect, a line segment lying on one line can be rotated about the point of intersection until it lies on the second line, and that successive such rotations about different points can be added to form a single rotation, in this case a rotation through half a circle, i.e. 180 degrees.

Inspired by thinking about Mary Pardoe's proof I once began to investigate a variant of Euclidean geometry (called "P-Geometry", in recognition of her contribution) in which the parallel postulate plays no role, but in which any line segment in a plane can be moved around in the plane, maintaining its length, and if such a motion includes a rotation, the angle of rotation has a

numerical measure. An incomplete discussion of P-Geometry is here (still work in progress):
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/p-geometry.html

Alternatives to Euclid's Postulates: Trisecting an angle in P-geometry

In 2013 I stumbled across information about Origami geometry and the fact that in Origami geometry there is a construction to trisect an arbitrary angle. The proof requires reasoning about 3-D structures produced by folding a planar surface: which is what Origami is all about. As a result it is not easy to follow.

More recently I learnt that a natural extension to Euclidean geometry known as the "Neusis" construction (discussed further in the Appendix below, possibly discovered by Archimedes) is a natural extension of Euclidean geometry that allows any angle to be trisected as shown below.

The Neusis extension is very closely related to P-geometry, insofar as both allow line-segments to be translated and rotated subject to constraints.

The Neusis construction allows a straight edge with two marks a fixed distance apart, to be rotated and translated subject to various constraints. Although this construction was apparently rejected as an extension to Euclidean geometry it is hard to see any justification for the rejection. Euclid had already allowed a pair of compasses to be moved in various ways while the end points remained a fixed distance apart, so why not allow a marked line segment to be moved while its marks remain a fixed distance apart?
A proof that the Neusis construction works is presented in:
    http://en.wikipedia.org/wiki/Angle_trisection#With_a_marked_ruler
and in This article.

However, since the proof is not easy to follow if one is unfamiliar with such reasoning, I offer a slightly clearer (I hope) presentation here, showing how it uses the feature of P-geometry that allows a fixed length line segment (the segment shown in blue, with two arrow heads) to be moved around in a plane, subject to certain constraints, changing its location and its orientation, but not its length.
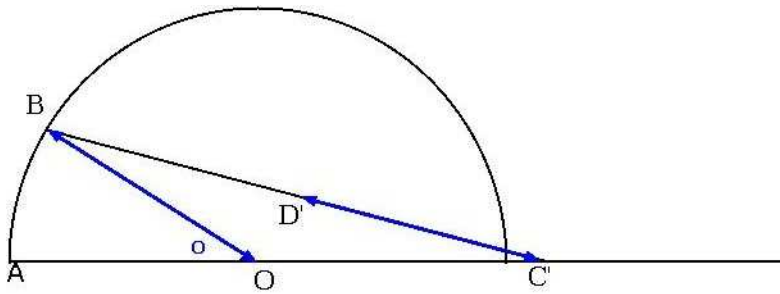
The proof is in three stages. For some readers the middle stage will be redundant. Like the Pardoe proof of the triangle sum theorem, this proof is best thought of as using not a static diagram, but a process in which a part of the diagram moves subject to constraints.
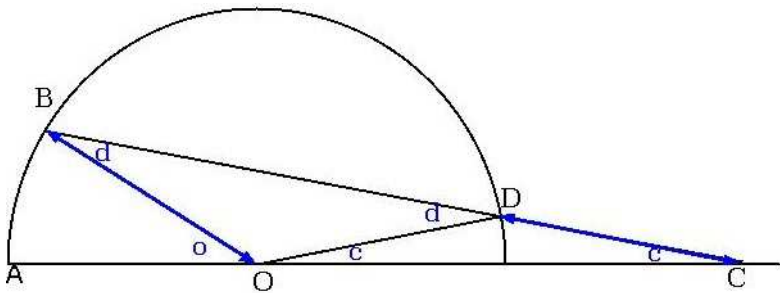

Stage 1:
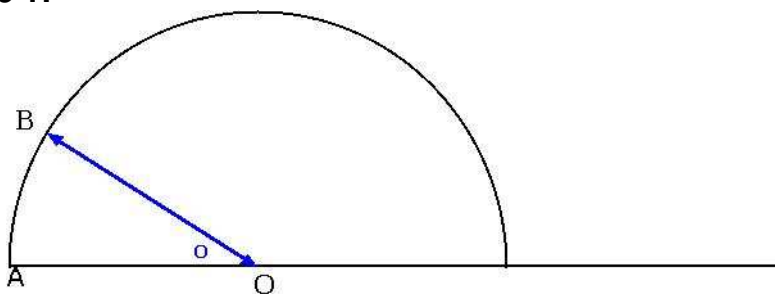
Stage 2:

_____

Stage 3:

-- Description:
The angle we wish to trisect is angle AOB, marked as "o", in Stage 1, and we shall show how to construct the angles marked "c" in Stage 3, so that they are exactly one third of "o".

The distances OA and OB are the same because A and B are on a circle drawn with centre O. The diameter of the circle through A and O is extended beyond the circle on the right. We now wish to create a copy of segment OB, which is a radius of the circle, and move it by sliding and rotating, to get a new configuration. That segment is shown in blue, as a double-headed arrow.

A copy of that blue segment slides to the right, with its lower end always on the diameter of the circle, if necessary extended beyond the circle on the right, and its other end always pointing at its original location, B on the circle. So the sliding process has three constraints:
1. The right end of the moving segment is always on the diameter through A, possibly extended to the right.

2. The line containing the moving segment always passes through location B, on the circle.

3. The length of the moving segment is constant.
The motion ends when the left edge of the moving segment is on the circle, at D. The right end is then at C.

The drawings show the initial stage, an intermediate stage and the final stage In Stage 2, as the right end of the blue segment C' slides to the right, the position of the left end D' keeps changing so that the distance C'D' remains equal to OB, and the orientation of the segment changes so that its extension always goes through point B, as shown in stages 2 and 3.

After the blue line has moved to the intermediate position C'D' shown in Stage 2, it continues sliding to the final destination CD shown in Stage 3, maintaining its length, and maintaining its relationship with point B.

The final destination is defined as the point at which the left end of the blue arrow lies on the circumference of the original circle, and its extension passes through B.

-- Proof that Angle 'c' is one third of Angle 'o':

Stage 3 repeated here for convenience of the reader:
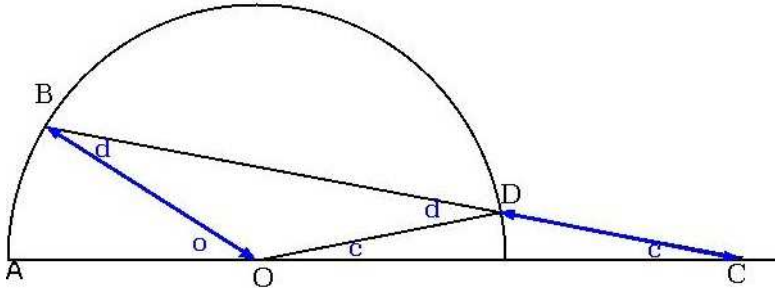
OB, OD, and CD must all be the same length. OD is the same length as OB because O is the centre of the circle on which both B and D lie. CD is the same length as OB because we kept moving the blue segment (from OB to CD) while keeping its length fixed, though the positions of the ends changed, and its orientation also changed while it slid.

Consider the triangle ODC: it has two sides the same length (OD, and CD) so the corresponding angles COD and OCD, are equal, both marked 'c'. It follows that the angle ODB, marked as having angle 'd', which is external to triangle COD, is the sum of the two opposite angles, OCD and COD, i.e.

```
    d = c + c
```

Now consider the triangle ODB. The sides OD and OB are the same length (radii of the circle), therefore the triangle is isosceles and angle ODB, previously marked 'd' must be the same as angle OBD, now also marked 'd'.

Now consider triangle OBC. Two of its angles, OCB and OBC have been marked as having sizes c and d respectively. They must sum to the external angle of that triangle, namely angle AOB, our starting angle, marked 'o'.

Therefore

```
    o = c + d
but
    d = c + c

therefore
    o = c + c + c
```

Therefore **c** is one third of **o**, so we have trisected angle AOB. The angles marked 'c', i.e. angles OCD and COD, are exactly one third of AOB.
QED

NOTE:
This construction started with an acute angle, AOB. If B had been moved round towards where D is, the angle would have been obtuse. The construction would then have to be different. Explaining why it would have to be different, and how to change it is left as an exercise for the reader.

Note added 20 Jul 2015:
How can we be sure that, provided that the initial angle AOB is less than a right angle (less than 90 degrees), the described constrained motion of the line segment will eventually reach the stopping state, i.e. left end of the segment on the circle, the right end on the line through A and O, and the extension of the line passing through point B?

It's "obvious" .. but why? What are the implications for P-geometry?

If the starting angle AOB is greater than a right angle the construction has to be changed. How to do this is left as an exercise.

NOTE Added 28 Feb 2020
I have just discovered that this construction is demonstrated nicely on the "cut-the-knot" Geogebra site here: http://www.cut-the-knot.org/pythagoras/archi.shtml
(However there's a minor bug, revealed by moving C too close to O! I've reported it, and it may be fixed.)

Implications for Philosophy of Mathematics

Immanuel Kant claimed, in his *Critique of Pure Reason* (1781) that mathematics includes knowledge that is synthetic, non-empirical, and necessarily true. Frege thought Kant was right about geometry, but wrong about arithmetic. He tried to show that arithmetic can be reduced to logic plus definitions, using logical deduction for all reasoning. That is, he tried to show that all the truths of arithmetic could be derived using axioms, definitions, and inference rules that are purely logical. But he felt that Kant was right about geometrical knowledge: it could not be derived from logic alone.

When David Hilbert published a logical axiomatisation of Euclidean geometry, Frege objected (roughly) that the axiomatisation was not about geometry, but about structures in another branch of (formalised) mathematics that happened to be isomorphic with at least a part of geometry.

For reasons that are not clear to me, Frege thought that his formalisation of arithmetic is not open to the same charge of "changing the subject". I think it is, as argued in my 1962 Oxford DPhil thesis defending Immanuel Kant's philosophy of mathematics.

> **NOTE (19 Jun 2016):**
> The thesis repeatedly used the impossibility of trisecting an angle as an example of a mathematical discovery illustrating Kant's claims. In principle the precise wording should now be revised, to allow for the possibility of adding the 'neusis' construction to Euclidean geometry.
> Thanks to efforts of a former student, Luc Beaudoin, the thesis (originally scanned by the Bodleian library in 2007) has now been retyped, and is available fully digitised, with searchable text, here:
> http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-1962

Two comments:

1. It is clear that humans are capable of creating, studying and using formal axiomatic systems and, if a system is consistent (and not too complex), discovering which mathematical theorems are true in it, i.e. they follow logically from the axioms. But if that process is inspired by a previously well-established branch of mathematics, the question arises whether the newly

created formal system using only logical reasoning simply defines a new portion of mathematics, with structural relationships to the old branch, or whether it articulates defining characteristics of the pre-existing branch, without changing the subject.

It seems that Frege, in effect, argued that Hilbert's axiomatisation of Geometry had changed the subject: instead of referring to geometrical structures and relationships the new axioms referred only to logical structures and relationships.

But Frege did not think this charge of changing the subject also applied to his own axiomatisation of arithmetic. My view was that Frege (like Kant) was right about geometry (i.e. it can be modelled logically but the original subject is not logic).

However, Frege was wrong about arithmetic being reduced to logic, since any logical formalisation of arithmetic, like any formalisation of geometry can be assessed according to how well it characterises what existed before the formalisation. For that, we need a good non-logical (or pre-logical) characterisation of arithmetical knowledge, comparable to the pre-logical characterisation of topological and geometric knowledge of spatial structures and processes, based on diagrams and imagined or experienced spatial configurations and processes. This topic is discussed further below.

2. It is not clear that there is only one pre-logical geometry. Although Euclidean geometry was, for a time, treated as having a special status, we now know that different variants of geometry are possible. The variants can differ in their degree of detail: e.g. projective geometry is less detailed than Euclidean geometry. They can also differ in the kind of detail. In particular, Euclidean geometry and P-geometry are different in ways mentioned above, including having different proofs of the same theorem, and having different theorems. If these versions of geometry are formalised using logic they must have different axioms and different valid proofs. The formalisation in logical terms leaves open the question how the different systems of concepts, different proofs, and different theorems arise. Where does knowledge about them come from originally?

An answer to that will need to be developed as part of the Meta-Morphogenesis project. I expect the answer to show how, as a result of the achievements of biological evolution, human brains, after appropriate stages of development after birth, have the potential to support non-empirical discoveries about spatial structures and processes, based on proofs exploring the scopes and limits of those structures and processes.

We can also construct models of those spatial structures and processes in arithmetic, as Descartes showed in his arithmetisation of geometry. But those models are new mathematical structures: they don't explain what the original knowledge was about. Likewise, when a mathematician produces a logical axiomatisation of Euclidean (or any other kind of) geometry, the original subject matter of geometry is clearly different from knowledge about a collection of logical axioms, definitions, and inference rules. Euclid, and Euclid's forebears knew nothing about the latter: since formalised logic, begun in a limited way by Aristotle, did not reach maturity till the 19th and 20th centuries. It was not available to the mathematicians who originally discovered and explored geometry.

The ancient Egyptians, Greeks and others who discovered how much can be done with a straight edge and compasses, had a type of insight into spatial structures and processes which enabled them to discover previously unknown properties of spatial structures. The mechanisms used in those discoveries must have been products of biological evolution, and probably overlapped with

spatial perception, planning, reasoning and action-control capabilities of other animals, even though those non-human animals, like very young humans (and poorly educated adult humans), lack the ability to reflect on, think about, and talk about the perceptual abilities they use expertly.

When someone later on discovered Origami geometry, perhaps as a result of playing with ways of folding paper (either using real paper or imagined paper) or came to think of geometry as allowing line segments to be moved and rotated in a plane, as in P-geometry, it is clear that this was a significant discovery that **extended** what had previously been known about Euclidean geometry, because the extended version does, and the earlier version does not, include a general procedure for trisecting angles, described above.

The possibility of that extension was already there in the structure of the space we perceive and act in, before humans noticed and discussed it. If those points are accepted, then that will provide support for the project of trying to understand the nature and content of geometric knowledge, and how it evolved, without treating that as merely an investigation of how human mathematicians discovered and explored a new logical structure. Moreover, it will need not have anything specific to do with *human* mathematicians since the same considerations will apply to any sort of mathematician able to think about the same topics -- waiting to be discovered before humans existed.

When Mary Pardoe first discovered her proof of the triangle sum theorem, she was not merely discovering that a new purely logical inference was possible starting from axioms she had previously been using, and neither was she merely exploring possible ways of assembling consistent sets of axioms and exploring the consequences.

Rather, while teaching mathematics she noticed for the first time a (pre-existing) feature of the realm of spatial structures and processes she was already familiar with, a feature that other mathematicians had not used in proving the triangle sum theorem. However, they had noticed that the ability to slide a marked straight edge around in the plane did allow constructions that were previously impossible in Euclidean geometry, including the above construction to trisect an angle.

It seems that that aspect of Euclidean geometry has not yet been studied in very great depth, since it if had been, it is likely that there would have been many reported discoveries, and re-discoveries, of the P-geometry proof of the triangle sum theorem. However an extension to standard Euclidean geometry closely related to P-geometry described in The Appendix, below had been discovered by Archimedes and used to trisect an angle.

What sort of discovery is involved in p-geometry?

Although the discovery may have been triggered by some mathematical experience ('awakened by experience' as Kant put it?) it was not an empirical discovery, derived from experience, and therefore liable to be refuted by future experiences. It shares that with the mathematical discoveries in Euclid. It is important to distinguish refuting the claim that something is a theorem in Euclidean geometry and refuting the claim that our physical space is Euclidean.

Did Einstein's theory of general relativity and the observations made during the 1919 solar eclipse (and more reliable observations later on) prove that geometry is empirical as often claimed?

No. The fact that Euclid's parallel axiom was found empirically not to fit large scale physical space, does not turn it into a mere empirical generalisation. There is a realm of mathematical reality, including non-Euclidean geometries, that had been discovered and explored before Einstein, by considering the implications of removing the parallel axiom from Euclid's axioms. That leaves a set of features (summarised in a subset of Euclid's axioms) that jointly specify a "core" type of space that does not intrinsically determine what happens if lines are extended indefinitely. Work by mathematicians before Einstein had shown that that core type of space could be extended in different ways that led to several distinct, interestingly different, sub-types of space, namely Euclidean space, with the parallel axiom, elliptical space (in which there are no parallel pairs of lines), and hyperbolic space in which a sort of parallelism has infinitely many sub-cases. For more details see:
http://en.wikipedia.org/wiki/Non-Euclidean_geometry

Within the framework of this expanded mathematical theory specifying a variety of types of space, with interestingly different theorems, there is indeed an empirical question about which type characterises the physical space we inhabit. That empirical question was answered theoretically by Einstein and his answer was confirmed empirically, initially by the observed displacement of stars during a solar eclipse, and later by many other measurements. But the discoveries of the non-Euclidean geometries, like discoveries of prime numbers and different sorts of equations, were non-empirical mathematical discoveries.

The non-empirical mathematical discovery preceding the empirical confirmation is that the type of geometry discussed by Euclid has a subset of features that can be expanded in different ways, one of which produces Euclidean geometry, while the other expansions produce non-Euclidean spaces. One of the reasons why this is mathematically interesting, is that each of the types of space raises interesting new questions, about which non-trivial new theorems can be proved. (When there are practical applications, such as predicting planetary motions, that adds to the importance of the mathematics.)

The unexpected empirical discovery by physicists was that one of the non-Euclidean alternatives best fits our physical space, although the deviations from Euclidean geometry are not normally noticeable. All this suggests that most of Euclidean geometry, without the parallel axiom, characterises physical space: that is an empirical claim that may prove to be false if there are regions of physical space (possibly on microscopic scales) where more of the Euclidean axioms break down.

What does not seem to have been widely appreciated is that Euclidean geometry with or without the parallel axiom can be extended to form P-geometry, as discussed above. So the Euclidean axioms, like the axioms of projective geometry, and the axioms (not yet fully formalised) of P-geometry, permitting translation and rotation of fixed-length line segments, characterise geometrically possible spaces, not merely logically possible sets of axioms. It may turn out that the defining features of P-geometry entail the parallel axiom. That would be another mathematical discovery. In that case, another possibility is that Euclidean geometry is a subset of P-geometry in the sense that all the geometric constructions possible in Euclidean geometry are also possible in P-geometry, but not vice-versa.

So perhaps the only theorems of Euclidean geometry that will turn out false in P-geometry are theorems of the form "no construction of type X is possible", e.g. "no construction guaranteed to trisect an angle is possible".

Poincaré on non-Euclidean geometry

There is a discussion closely related to this topic, by the great mathematician Henri Poincaré originally in a collection published in French in 1902. An English translation was published in 1905. One of the articles, on Non-Euclidean geometries, has been made available here:
http://www-history.mcs.st-and.ac.uk/Extras/Poincare_non-Euclidean.html

He implicitly, and very briefly, discusses something like P-geometry (without naming it) when he points out that there are different geometries: some of them, but not all, allow motion of a figure all the lines of which remain of a constant length. P-geometry would be an example of that subset. It is possible that he had previously learnt about the Neusis construction discussed in the Appendix, or had reinvented something like it.

After discussing the existence of various geometries with different collections of axioms and theorems, he writes:
Are the axioms implicitly enunciated in our text-books the only foundation of geometry? We may be assured of the contrary when we see that, when they are abandoned one after another, there are still left standing some propositions which are common to the geometries of Euclid, Lobachevsky, and Riemann. These propositions must be based on premisses that geometers admit without enunciation. It is interesting to try and extract them from the classical proofs.
He mentions that some but not all geometries have the properties of our P-geometry:
The possibility of the motion of an invariable figure is not a self-evident truth.
Then later:
If geometry were an experimental science, it would not be an exact science. It would be subjected to continual revision. Nay, it would from that day forth be proved to be erroneous, for we know that no rigorously invariable solid exists. The geometrical axioms are therefore neither synthetic apriori intuitions nor experimental facts. They are conventions. Our choice among all possible conventions is guided by experimental facts; but it remains free, and is only limited by the necessity of avoiding every contradiction, and thus it is that postulates may remain rigorously true even when the experimental laws which have determined their adoption are only approximate. In other words, the axioms of geometry (I do not speak of those of arithmetic) are only definitions in disguise. What, then, are we to think of the question: Is Euclidean geometry true? It has no meaning. We might as well ask if the metric system is true, and if the old weights and measures are false; if Cartesian co-ordinates are true and polar coordinates false. One geometry cannot be more true than another; it can only be more convenient. Now, Euclidean geometry is, and will remain, the most convenient: 1st, because it is the simplest, and it is not so only because of our mental habits or because of the kind of direct intuition that we have of Euclidean space; it is the simplest in itself, just as a polynomial of the first degree is simpler than a polynomial of the second degree; 2nd, because it sufficiently agrees with the properties of natural solids, those bodies which we can compare and measure by means of our senses.
(I don't know what he would have thought about the general theory of relativity, had he lived long enough to encounter it.)

I think Poincaré was nearly right, but not quite. The ability to select, characterise and label a particular type of geometry requires more than just the formulation of a definition. It presupposed mathematical discoveries concerning possible types of space, and in particular alternatives to Euclidean space. Compare a mythical community who had first discovered that a closed figure bounded by straight lines is possible when they first drew or thought of a triangle, or found a naturally existing triangular shape. They could then have started exploring properties of triangles,

teaching that as a branch of mathematics, one of whose axioms was something like "Every shape has three sides" -- the axiom of triangularity.

Later someone might discover a quadrilateral, or a pentagon, and then complain that the axiom of Triangularity is just an arbitrary definition, which could just as well be replaced by "Every figure has three or four sides". But further research would show that neither triangles nor quadrilaterals, nor their union, exhausted the space of possible closed figures. (Piaget's posthumously published book on *Possibility* shows similar stages of development in considering possibilities in children between the ages of about three and thirteen. A deep challenge for developmental psychology, neuroscience and AI is to explain and model what happens during such development.)

At that point attention could switch to finding out what sorts of closed shape can exist and then different branches of shape research could develop studying polygonal shapes with different numbers of sides, convex shapes, non-convex shapes, shapes bounded by curves, etc.

Definitions, perhaps labelled as axioms, could be used to introduce new labels for particular types of shape under investigation. But far from being trivial new abbreviations for arbitrary combinations of features (e.g. "A quachelor is an unmarried male who has proposed marriage unsuccessfully to four different women"), the new axioms, in combination with more general axioms characterising a space of possibilities, usefully delimit non-trivial sub-fields of mathematical investigation. Moreover, the implicit *existential* claim that the sub-field has instances, or can have instances, ensures that such new axioms, whether regarded as definitions or not, express significant content.

(The APPENDIX below contains further discussion on how to view the axioms describing possible constructions and their properties, in different geometries.)

"Geometries" of symbolic forms

It is worth noting that when we use a formal language and perform syntactic operations to check that a sentence in that language is well formed, and to decide whether a particular sentence is derivable from specified axioms and inference methods, then we are studying or using a domain of formulae, i.e. linear discrete strings conforming to a set of syntactic rules (a grammar). But for that study we need a sort of geometry (or at least topology) in which there are sequences of locations that can be occupied or not occupied, and if occupied the contents of the locations must take various forms specified by the grammar of the language.

Checking well formedness and checking derivability of a formula within a particular system require abilities to reason about a class of linear structures, e.g. to infer what the consequence will be of substituting a particular symbol or structure for instances of a variable in the formula. In other words, doing mathematics as a formalist requires specific mathematical abilities that are as sophisticated as the abilities required to study geometry. But they are different abilities. This point was made in my 1962 thesis and has also been made by others (Hilbert?). It demonstrates that even a formalist philosophy of mathematics cannot avoid the task of explaining what sorts of cognitive competences are involved in doing mathematics and understanding enough about a domain of structures to be able to cope with a novel one.

Moreover, although a Turing machine or modern computer can operate on symbolic structures that are treated as composed of symbolic structures with no inner structure (e.g. bits), human mathematicians using logical formalisms have located their symbols in the same perceptual spaces as they use for the rest of their life. And the ability to do that needs to be analysed and explained.

From my point of view, the ability to do logic, algebra or manipulate symbolic structures is just another mathematical ability that somehow arose within an evolutionary trajectory, and needs to be explained as part of a general explanation of the nature of mathematical abilities. (Some of those abilities have been replicated in AI systems, e.g. theorem provers, parsers, etc. but so far only with narrowly focused meta-linguistic capabilities. I doubt that any current AI system could have discovered the natural numbers and invented the notations we now use for representing numbers. (Doug Lenat's PhD thesis described a system *AM* that had a subset of the required capabilities. http://en.wikipedia.org/wiki/Automated_Mathematician)

My attempts to answer some of these questions in more detail, in a long-term project investigating Biological Evolutionary Foundations of mathematics (still in its early phases) are presented in a paper on construction kits of various types (work in progress) and their roles in evolution and individual development.

The existence of these interesting, deep, mathematically explorable abstract spaces is not a trivial theorem of logic. Rather logic is one of the spaces.

Natural numbers and one-one correspondence

An argument can be presented showing why arithmetic of natural numbers is not a branch of logic, as Frege and Russell thought, but a branch of mathematics concerned with properties of one to one correspondences. For example, the existence of prime numbers amounts to the impossibility of putting certain collections of objects into a one to one correspondence with an **M by N** regular array of objects, where both **M** and **N** are numbers greater than 1. More generally, someone playing with one-one correspondences could discover that in some cases a collection of objects can be put into a one to one correspondence with another collection that is composed of two or more sub-collections each of which is in one-one correspondence with each of the others. Those are the collections that are "factorizable".

There are other collections that cannot be put into a one to one correspondence with the union of two or more equinumerous collections. Proving that that is impossible for collections of 7, 13, 19 objects, for example, is at first sight harder if the objects are not all of the same size and shape. But that can be remedied by starting with rectangular blocks all the same size, then showing that if S and T are in a one-one correspondence and S can be factored, then so can T, then showing that any finite collection of objects can be put into a one-one correspondence with a collection of cubes, or squares, all the same size.

Examples relating to both arithmetic and geometry (including topology) are presented in this incomplete discussion of "toddler theorems":
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/toddler-theorems.html
Both are parts of the Meta-Morphogenesis project:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-morphogenesis.html

How the concept of 'primeness' can be discovered spatially is discussed here:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/toddler-theorems.html#primes

NOTE (Thanks):
Mary Pardoe, who originally provoked this line of thought in me, is actively involved in mathematical education, as indicated by her Twitter page: https://twitter.com/pardoemary and her contributions to the National Center for Excellence in the Teaching of Mathematics (NCETM), illustrated here:
https://storify.com/ncetm/how-can-we-use-the-element-of-surprise-in-maths-le

A different proof of the triangle sum theorem, using external rotation angles (complementary to the internal angles of the triangle) was discovered by Kay Hughes (QinetiQ) in 2013, while we were informally discussing mathematics education (at the dinner of my wife's Orienteering club), as reported here:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/triangle-sum.html#kay
It took her about five minutes of thinking, without use of pencil, paper, etc.

APPENDIX: Note on the Neusis construction
(Added 1 Mar 2015)

The Neusis construction is described and discussed on Wikipedia here:
http://en.wikipedia.org/wiki/Neusis_construction

It is normally described in terms of extending Euclidean planar geometry by allowing use of a rigid but movable straight-edge or straight line S (e.g. a ruler) with one or more locations marked on the S, one of which is a "Tracker point" T which is fixed on S but able to move in the plane, a straight or curved line "Guide line" G in the plane which is used to "guide" the motion of S, a "pivot point" P fixed in the plane with which S must remain in contact as it slides in the plane, and another straight or curved line the "Catch line", which can be used to express a constraint on the motion of S.

The motion has two main constraints:

1. The "Tracker point" T (e.g. the right end of the blue arrow in the construction above, which starts at location O and then moves to the right) moves along the guide line G (e.g. the extended line AB above).
2. The "Pivot point" P is fixed in the plane (e.g. location B above), but can slide along the ruler as the ruler is moved according to the first constraint.

In our example the Catch line is the circle with centre O and passing through A and B. Its function in this case is to determine when the motion of S should end, namely when the distance between tracker point T and the intersection of S with the catch line equals the radius of the circle, i.e. when CD=OA.

If the guide line G is straight and the tracker and pivot points T and P both lie on G, then the straight-edge S will move on G and its orientation will not change. In all other cases as T moves, S will rotate around the pivot point P.

If G is a circle with centre at P, then the point on S that coincides with P will not move on S. In all other cases, as T moves along G (T tracks G), P will slide along S.

Note: According to the Wikipedia page, the point P is referred to as the "Pole", the guide line G is referred to as the "directrix", and the length CD to be reached between T and the Catch line is referred to as the "diastema".

It is possible to think of this construction (and all constructions in Euclidean geometry) in at least two different ways, namely either as assuming a physical world in which there are rigid movable objects, such as straight edges, marked rulers, compasses (or rotatable fixed-length line segments with one end fixed), or as assuming only structures and processes in a space of possible shapes and motions, which may or may not be occupied by objects with physical properties. These are examples of physical and abstract construction kits, discussed in a separate paper.

This is analogous to the difference between regarding a logical notation as concerned with physical formulae written on physical surfaces using physical coloured material, and thinking of logical notation as dealing with abstract structures and patterns that can be instantiated on physical surfaces but have properties about which we can reason without assuming any particular physical instantiation. So, for example, the formulation of the law of excluded middle in propositional calculus is not usually taken to assume that logic is concerned with ink marks on paper, or chalk marks on slate, or, for younger readers, patterns of illumination on a physical screen.

However in the logical case, besides the ambiguity between a space of physical constructions of patterns in space and abstract constructions in a space of abstract structures, there are also differences between statements about those structures themselves (e.g. which are derivable from which in the relevant construction kit) and statements about their semantic content, which in the case of propositional logic might be expressed in terms of truth tables, or more generally truth tables plus semantic rules specifying how to assign truth values to non-logical assertions (e.g. "Tom is taller than Fred").

I think some of the more obscure features of Kant's philosophy of mathematics are concerned with these distinctions between discoveries about the world and discoveries about the properties of various construction kits. Philosophers and mathematicians who reject some of Kant's claims about geometrical reasoning seem to me implicitly to accept similar claims in relation to logical reasoning. (The fact that logical rules and formulae, along with arithmetical theorems and proofs, have been given such a special status since around 1900 will probably turn out to be merely a temporary phase in the history of mathematics.)

In arithmetic based on a set of axioms and rules of inference, the status of axioms or theorems as expressing *necessary* truths can by undermined by considering the difference between ordinary arithmetic (Peano arithmetic) and modular arithmetic, e.g. arithmetic modulo 5. In Peano arithmetic 0 is not the successor of anything, whereas in modular arithmetic it is. Moreover in Peano arithmetic the set of numbers is infinite, whereas in arithmetic modulo 5 it is finite. Is there any reason why we should not regard modulo arithmetics and non-euclidean geometries as comparable mathematical discoveries: namely both turn out to be products constructable within certain mathematical construction kits, one spatial the other more abstract.

However the arithmetic that is based on properties of one-one correspondences comes from a different construction kit. I leave it to the reader to work out whether and how modular arithmetic can turn up in that context.

I suspect full clarity on this issue will not be possible until we have a much deeper understanding of the evolution and functions of various kinds of construction kits, including concrete, abstract and mixed construction kits.

A construction attributed (almost wordlessly!) to Rufus Isaacs in Roger Nelsen (1993). The construction is depicted thus (by Nelsen):

Figure Trisect
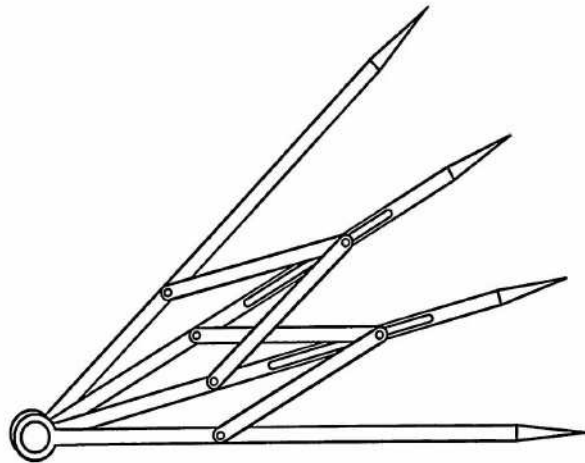
Apparatus for trisection of an arbitrary angle.

Added 25 Feb 2020
For a more detailed discussion of essentially the same mechanism see:
http://www.iri.upc.edu/people/porta/Soft/CuikSuite2-Doc/html/dir_d4321ff35d21dae68cd3d54c562c7078.html

---

Wittgenstein on trisection

Juliet Floyd 1995, discusses the use made by Wittgenstein of the impossibility of trisecting an angle in Euclidean geometry. Although she mentions the Neusis construction, it appears that Wittgenstein did not know of it. Insofar as learning that the Neusis construction is possible and has the consequences described above, is a mathematical discovery, not an empirical one it seems to be comparable to the discovery of alternatives to the parallel axiom in non-euclidean geometries. Floyd does not really discuss the status of that discovery.

JUMP TO CONTENTS

REFERENCES AND LINKS

* Juliet Floyd, 1995, On saying what you really want to say: Wittgenstein, Gödel, and the Trisection of the Angle, in *From Dedekind to Gödel: Essays on the Development of the Foundations of Mathematics,* Ed. Jaakko Hintikka, Springer, Dordrecht,Netherlands, pp. 373--425,

http://philpapers.org/archive/FLOOSW.pdf

- Freksa, Barkowsky, Falomir and van de Ven, (2019), Geometric problem solving with strings and pins, *Spatial Cognition & Computation,*
https://doi.org/10.1080/13875868.2018.1531415

- Roger B. Nelsen, (1993), *Proofs without words: Exercises in Visual Thinking,* Mathematical Association of America, Washington DC.
This is one of several online versions of this wonderful collection:
https://is.muni.cz/el/1441/podzim2013/MA2MP_SMR2/um/Nelsen--Proofs_without_Words.pdf

- Much of Jean Piaget's work is also relevant, especially his last two (closely related) books written with his collaborators:
Possibility and Necessity
Vol 1. The role of possibility in cognitive development (1981)
Vol 2. The role of necessity in cognitive development (1983)
Tr. by Helga Feider from French in 1987

- http://www-history.mcs.st-and.ac.uk/history/HistTopics/Trisecting_an_angle.html
Trisecting an angle, by John J O'Connor and Edmund F Robertson, in *MacTutor History of Mathematics archive*
http://www-history.mcs.st-and.ac.uk/history/
School of Mathematics and Statistics University of St Andrews, Scotland

- http://www.cs.bham.ac.uk/research/projects/cogaff/misc/construction-kits.html
Aaron Sloman
Construction kits required for biological evolution
(Including evolution of minds and mathematical abilities.)
The scientific/metaphysical explanatory role of construction kits

- http://www.cs.bham.ac.uk/research/projects/cogaff/misc/impossible.html
Some (Possibly) New Considerations Regarding Impossible Objects
Their significance for mathematical cognition, and current serious limitations of AI vision systems.

JUMP TO CONTENTS

Maintained by Aaron Sloman
School of Computer Science
The University of Birmingham

--