On Mathematical Notation

Mark Colyvan
Department of Philosophy, University of Sydney
A14 Main Quadrangle Sydney, NSW, 2006, Australia
mark.colyvan@sydney.edu.au

Abstract

There is something right about the view of mathematics as "the language of science". Thinking of mathematics as a language is useful in appreciating the significance of, and the difficulties encountered arriving at, a good notational system. Good notation is far from trivial. But thinking of mathematics as *merely* language, is to ignore the other roles mathematics can play in science. I will consider the role good notation can play in prompting new ideas and new developments in mathematics and science. Notation may even be thought to contribute to mathematical explanations.

Keywords: Mathematical Notation, Explanation, Philosophy of Mathematics.

1 The language of science

One often hears the claim that mathematics is "the language of science". This, I take it, is meant as a compliment to mathematics. The point of the slogan is to emphasise that a great deal of science — especially physics, but many other branches as well — is typically highly mathematical and could not even be formulated without mathematics.

It is worth taking seriously the idea that mathematics is the language of science, for there is undoubtedly something right about it. But in doing so we need to guard against a temptation to view mathematics as *merely* the language of science — a medium for expressing scientific ideas but not contributing to the scientific enterprise in any other way. We need to remember that mathematics plays a variety of roles in many diverse branches of science. In this paper I will consider some of these roles and show how the language of mathematics (i.e., mathematical notation) contributes to

Talk at the conference "Philosophy, Mathematics, Linguistics: Aspects of Interaction 2012" (PhML-2012), Euler International Mathematical Institute, May 22–25, 2012.

these roles. In particular, I will show how good, clear, and economical mathematical notation does not only assist in the communication of ideas, it also facilitates new ideas and understanding.

It is interesting to note that mathematicians and historians of mathematics have long recognised the importance of good notation in mathematics.[1] But since the fall from grace of formalism as a philosophy of mathematics, mathematical notation has received little attention in philosophical circles.[2] Formalists saw the subject of mathematics as nothing but the notation and its manipulation. While Gödel's incompleteness theorems were supposed to have put paid to that view — at least to Hilbert's well-developed version of formalism — it is worth reflecting upon what was right about formalism and its attention to mathematical notation. Notation *is* very important to mathematics and a great deal of mathematical activity is concerned with the manipulation of symbols. Formalist went too far in suggesting that mathematics was nothing more than notation and its manipulation, but let's not forget the formalists' insight.

2 The value of good notation

There are a number of well-known examples used to illustrate the value of notation in mathematics. Chief among these examples is the power of Arabic notation for the natural numbers over Roman notation [4]. Some of the virtues of good notation include transparency (the notation reminds the reader of what it stands for), economy (the notation is not to unwieldily and is able to efficiently convey often complex ideas), and calculational power (the notation readily lends itself to various mathematical manipulations.[3] There is no doubt that these virtues are important, but I am interested to explore examples that suggest that there is even more to good notation.

I will consider examples where the notation plays a part in extending mathematical theory and plays a role in delivering explanations. My examples and discussion will also go beyond what might be thought to be relevant to a natural (but narrow) construal of the topic of mathematical notation. I am interested in mathematical notation in a very general sense. Indeed, mathematical notation is a representational tool and, as such, has a great deal in common with other forms of mathematical

[1] For example, see J.E. Littlewood's interesting remarks about notation [16, p. 35]. See [5, 6], for the history of mathematical notation and [13] for a linguistic analysis.

[2] One notable exception is James Robert Brown [4] and, more recently, Catarina Dutilh Novaes [11]. Other than these, there are only a handful of philosophical papers on the topic (e.g., [10, 19]).

[3] For example, Arabic notation for the natural numbers lends itself to basic mathematical operations of addition, division and multiplication. Compare long division in Roman notation.

representation such as diagrams and constructions (e.g. set theoretic constructions of the natural numbers via, for example, von Neumann ordinals [12]). Although I won't pursue the more general topic of mathematical representations here, I hope to show, by way of my chosen examples, that the topic of mathematical notation is richer than one might initially think and has some interesting and natural connections with this more general topic of mathematical representation.

This first example illustrates how clever notation can reveal something mathematically interesting. Here we show that one can add infinity to the complex plane, to deliver the extended plane, without leading to trouble. This is done by a construction called *the stereographic projection*. But this construction can also be considered to be providing an alternative notation for the complex plane. Here's how it goes. Consider the complex plane with each point represented with the usual Cartesian 2-coordinates (x, y). Add a third z dimension. Represent this three space with the usual Cartesian 3-coordinates (x, y, z). Now consider a unit sphere with its center at the origin $(0, 0, 0)$: $x^2 + y^2 + z^2 = 1$. Take the line generated by joining a point in the the x–y plane with the north pole of the sphere $(0, 0, 1)$. In particular, we are interested in where this line intersect our unit sphere. For example, the line joining the point $(1, 0)$ in the x–y-plane and the north pole of our sphere intersects the sphere at $(1, 0, 0)$. The line joining the point $(0, 0)$ to the north pole, intersects the sphere at the south pole $(0, 0, -1)$. If we now identify each point of the Cartesian plane with the corresponding point of intersection of the sphere, we see that all the points inside the unit circle $x^2 + y^2 = 1$ are mapped to unique points on the part of the sphere below the x–y-plane, all the points on the unit circle get mapped to their 3-dimensional counterpart (i.e., (x, y) gets mapped to $(x, y, 0)$), and all the points outside the unit circle get mapped to unique points on the portion of the sphere above the x–y-plane. Moreover, the further from the origin the point is, the closer its corresponding point of intersection is to the north pole of the sphere. Via this conformal mapping, we have created an alternate representation for each point in the real plane.

What is useful about this construction, is that every point on the plane is represented by an ordered triple (points on the sphere) and every point of the sphere, except one — the north pole — has a corresponding point on the x–y-plane. But now we just add the north pole and stipulate that it is the point at infinity. The motivation for this should be plain to see. It should also be clear that the north pole represents infinity in all directions on the x–y-plane. Using this construction (or alternative notation) we have shown that it is sensible to talk of the extended real plane (i.e., the complex plane with a single point at infinity added). That this can be done is not at all apparent in the standard notation. We have found an interesting extension of our mathematics, and this extension was facilitated, at least in part, by the alternative notation employed in the stereographic projection.

3 Notation and explanation

So far we have seen that good notation can enjoy the virtues of transparency, economy, calculational power, and, as the last example illustrated, it can facilitate advances in mathematics. But let's push things a little further. Let's pursue the suggestion that good notation can make mathematical explanations more perspicuous.

Mathematicians talk of explanatory and non-explanatory proofs in mathematics [14]. The explanatory proofs tell you why the theorem in question holds, while the non-explanatory proofs merely tell you that it holds. The explanation is something delivered by the proof and, of course, there can be explanatory and non-explanatory proofs of the same theorem. Exactly what this explanation amounts to is hard to say. Indeed, it is hard to say, in general, what an explanation is. In empirical science and in everyday life, very often a request for an explanation is a request for details of the causal history. Why did the window break? Because it was hit by a rock. Why did the anode emit X-rays? Because it was bombarded by high energy electrons. In mathematics, however, whatever an explanation amounts to, it can't be anything to do with causal history. Explanations of why the Poincaré conjecture is true, cannot have anything to do with causal histories. Apart from anything else, if mathematical objects such as functions and derivatives exist, they do not have causal powers and they do not enter into causal networks. Causal accounts of explanation, whatever virtues they have elsewhere in science, simply miss the mark in mathematics. So, then, what is a mathematical explanation?

One promising line on mathematical explanation is that it is unification.[4] According to this account of explanation, we explain some phenomenon when we incorporate it into a wider class of phenomenon. To give a classic example: Newton's theory of gravitation unified planetary motion, terrestrial projectiles, and the tides. Moreover, in doing this it explained these phenomenon. It is important to note that on this view, Newtonian gravitation explains the tides, even though the reason for the tides is explained in terms of a somewhat mysterious force acting at a distance: gravitational attraction of massive bodies. Explanations do not need to bottom out in self-evident truths and they don't have to eliminate all mystery. In mathematics, we can see how this might work. Seeing connections between apparently unrelated parts of mathematics is indeed enlightening, and arguably explanatory. For example, Euler's formula, $e^{i\theta} = \cos\theta + i\sin\theta$ reveals deep connections between trigonometry

[4]See [15, 17] for more on this account of explanation.

and complex analysis (at least from the point of view of trigonometry). Representing systems of simultaneous equations in linear algebra enables results about when solutions can be found, in terms of the properties of the corresponding matrices (determinants, eigenvalues and so forth).

That's explanation within mathematics. Some have gone on to suggest that mathematics can explain empirical facts.[5] For example, Alan Baker [1] suggests that the reason North American cicadas have life cycles of 13 and 17 years involves facts about prime numbers. It turns out that having a life cycle of 17 years (or some other prime number in the vicinity) is a very effective way of avoiding predators, since only those predators with the same prime life cycles or annual predators will coincide with the cicadas at their most vulnerable stage. Other life cycles, 14 years, say, may coincide with predators with 14 year, 7 year, 2, year and 1 year cycles. Prime life cycles for such semelparous reproductive species increases survival rates. And a major part of the explanation of why this is so comes from number theory. This explanation is not purely mathematical — it also involves facts about predators and the the cicada life cycle — but it does have a mathematical component to the explanation. In other words, an explanation of the 13 and 17 year life cycles of North American cicadas would not be complete without an appeal to the fact that 13 and 17 are prime numbers.

Now we return to the main topic of this paper: mathematical notation. I will demonstrate how good mathematical notation can help engender explanations, both intra-mathematical explanations and extra-mathematical explanations.

3.1 Analytic geometry

Chief among René Descartes' many contributions to philosophy, mathematics, and science, is his idea of combining geometry and algebra to give rise to *analytic geometry*. The idea is so familiar these days that it is hard to fully appreciate just how brilliant this innovation was. We invoke a coordinate system for the plane, then two-dimensional geometric figures can be represented algebraically in terms of these coordinates. In the other direction, we can represent algebraic equations and inequalities geometrically. With this algebraic notation for geometric figures, we are able to use geometry to help visualise otherwise abstract algebraic problems and to use the rigorous methods of algebra to solve geometric problems.

For example, from the relevant geometry we can see why the polynomial $x^2 - 1$ has 2 real roots: because the corresponding parabola $y = x^2 - 1$ has its vertex below the x-axis. We can also see why the polynomial $x^2 + 1$ has no real roots:

[5] See, for example, [2, 8, 9, 18] for discussion of whether mathematics can explain empirical facts.

because the corresponding parabola $y = x^2 + 1$ sits entirely above the x axis.[6] It is, of course, possible to understand why $x^2 + 1$ has no real roots by considering the algebra alone, but with the connection to geometry in place, you can *see* why. This is just the tip of the iceberg. By invoking the power of analytic geometry we can visualise differentiation (as the function representing the slope of a tangent to a curve at a given point), integration as the area under a curve, and so on. These and many other applications are familiar. It is worth reflecting for a moment, however, on the important role the algebraic notation for geometric figures plays in all this. Indeed, all there is to analytic geometry is the merging of algebraic notation and geometry, and somehow both geometry and algebra are enriched by the merger.

One striking example of the power of algebraic methods in geometry is the Weierstrass function. It is tempting to believe that continuous functions can fail to be differentiable at only a countable number of points. That is, if a function is continuous on an interval, it may have at most a countable number of points at which it is not differentiable. Weierstrass demonstrated that this natural and intuitive suggestion is mistaken. He showed that there are continuous functions that are nowhere differentiable — they are nothing but cusps. The function Weierstrass originally produced was

$$f(x) = \sum_{n=0}^{\infty} a^n \cos(b^n \pi x),$$

where $0 < a < 1$, b is a positive odd integer and $ab > 1 + 3\pi/2$. Once you see the trick involved, you see that there are many other such functions. It is very hard to imagine discovering such functions without the power of algebraic notation at one's disposal. After all, pure geometric intuitions, if anything, suggest that are no such functions. But more importantly, the algebraic notation allows the explanation of why some continuous functions may fail to be differentiable on intervals.[7]

My suggestion in this and in the following example is that the notation helps us understand what is going on, and in this sense, helps engender explanations. By this I do not mean to suggest that the notation is the reason for the phenomenon in question; I just mean that the notation helps make the explanation accessible to us.[8]

[6] Although, of course, the fundamental theorem of algebra, guarantees that the polynomial in question has 2 complex roots: $\pm i$, where $i = \sqrt{-1}$.

[7] Of course, study of the Weierstass function is conducted as part of analysis. Indeed, that, in part, is my point here: geometric intuitions and visualisations will not deliver this function. To employ analysis algebraic notation must be available.

[8] We can distinguish two senses of explanation: one an objective sense and the other a psychological sense. For example, a scientific explanation might not be understandable to someone not versed in the relevant science. That does not stop it from being an explanation in the objective sense. But its failure to enlighten the person in question prevents it from being a psychological

Next we consider an example where algebraic notation is used in a more essential way. Again the explanation in question relies on the algebraic representation of geometry.

3.2 Squaring the circle

Squaring the circle is the well known problem of constructing a square of the same area as a given circle, using just a straightedge and compass.[9] That is, using only these two instruments we must construct a square of side $r\sqrt{\pi}$, where r is the radius of the circle in question and π is the ratio of the circumference to the diameter of a circle. It is well known that after centuries of attempts, the impossibility of this construction was finally proven in 1882 by Ferdinand von Lindemann. The proof, however, comes from abstract algebra, not geometry. A quick sketch of the connection here is worthwhile, because the algebraic notation is crucial to the proof in question [3, 39–48].

First, we catalogue the legitimate, basic ruler and straightedge constructions (drawing a line through two existing points, constructing a circle with center at one existing point and running through another existing point, and so on). We then provide notation for the basic geometric objects (lines, points, and arcs of circle) and note that we can represent these objects in the Cartesian plane, in the usual way. We then show that the permissible geometric constructions give rise to a small set of algebraic operations on line lengths: addition, subtraction, division, multiplication, and taking the square root. The idea here is that if we are given a unit line segment and two line segments of lengths, a and b, we can construct line segments of length ab, $a+b$, $a-b$, a/b and $\sqrt{a}$. What is crucial is that these are the *only* algebraic operations the geometric constructions licence. Note that what we have done is again forge a link between the geometric constructions and algebra. This allows us to apply algebraic methods to the problem. We have thus transformed the problem from one of geometry — that of constructing a given length, $\sqrt{\pi}$ — to one of abstract algebra, namely that of determining whether $\sqrt{\pi}$ can be obtained by successive applications of the algebraic operations just listed. In particular, the notation just described turns this into a problem of determining whether $\sqrt{\pi}$ can be the root of a polynomial with powers 0, 1, or an even integer, and with rational coefficients. This is where the Lindemann result comes in. Lindemann proved that

explanation for that particular person. It is this latter psychological notion of explanation we're concerned with here.

[9] This is one of three famous ancient geometric construction problems. The other two problems are trisecting an angle, and doubling the cube. The former is the problem of trisecting and arbitrary angle, the latter is that of creating a cube with twice the volume of a given cube.

π (and therefore $\sqrt{\pi}$) is transcendental. That is, π (and $\sqrt{\pi}$) is not the root of any polynomial with rational coefficients. The ruler and straight edge construction is thus impossible.

Although the fact that π is transcendental is the key to the impossibility result, it is important to see how the problem needed to be set up as an algebraic problem. This involved the introduction of algebraic notation for the geometric objects and operations, and noting that the geometric operations give rise to some familiar algebraic systems.[10] Again we see good mathematical notation playing a key role in delivering a mathematical explanation. But notice that the explanation goes beyond pure mathematics. We have also explained why all attempts to square the circle failed and why anyone who claims to have squared the circle is not taken seriously [7].[11]

4 Shakespeare's mistake

We have seen that good notation can enjoy the virtues of economy and calculational power, and can facilitate advances in mathematics. More controversially, good notation may also contribute to mathematical explanations and perhaps even explanations beyond mathematics. This already takes us well beyond the standard view (insofar as there is such a thing) of mathematical notation. The standard view I have in mind here, suggests that it's the mathematical objects that matter, not the notation we use for them: as Shakespeare put it in *Romeo and Juliet*:

> What's in a name? that which we call a rose
> By any other name would smell as sweet.

Following Shakespeare's lead, it might be tempting to suggest that a mathematical object by any other notation would be just as useful. But we have seen that this is not so. Sometimes the notation encode properties of the objects in question and in such cases, alternative notations would be less revealing.[12] Perhaps because of this encoding, good notation can keep track of distinctions we may not have initially noticed, force us to investigate such distinctions, and to see possibilities for

[10] It is worth noting that the transcendence of π was proved by analytical methods, whereas the impossibility of trisecting the angle, and the impossibility of doubling the cube were proved by algebraic methods. But in all three problems, the algebraic representation of the geometric constructions was central.

[11] See [4] for more examples of the power of mathematical notation.

[12] This is true of natural language as well: onomatopoeias reveal something about the sounds they name and other words such as "football" and "calculator" tell you about the objects they name.

future research not previously anticipated. In short, properties of the notation are important in mathematics. So Shakespeare was wrong, at least about mathematical notation, but I'd suggest he was wrong about natural language as well. A rose by any other name may well smell as sweet but sometimes the name chosen is important; smell isn't all we're interested in. In any case, it is far from clear that mathematics would be served equally well by alternative notations. Getting the notation right features prominently in mathematical practice. And there is a good reason for this: good notation does serious work in mathematics. And it is for this reason that mathematicians spend so much time developing and discussing notation.

Mathematics may well be the language of science but we must not underestimate the power of this language for both expressive purposes and in facilitating new ideas and explanations. Good mathematical notation, like any powerful language, helps us to think and to gain understanding.

Acknowledgements

Thanks to Jim Brown, Jeremy Butterfield, Jeff Pelletier, Graham Priest, Matthew Spillane, and Zach Weber for very helpful discussions on the topic of this paper and to Oleg Proserov and an anonymous referee for helpful comments on an earlier draft. I am also indebted to audiences at the 2009 Australasian Association of Philosophy Conference at the University of Melbourne, the 2011 European Philosophy of Science Association meeting at the University of Athens, and the 2012 "Philosophy, Mathematics, Linguistics: Aspects of Interaction" conference at the Euler International Mathematical Institute, St. Petersburg, Russia. Special thanks to Oleg Proserov for organising the St. Petersburg conference.

References

[1] Alan Baker. Are there genuine mathematical explanations of physical phenomena? *Mind*, 114:223–238, 2000.

[2] Robert W. Batterman. On the explanatory role of mathematics in empirical science. *The British Journal for the Philosophy of Science*, 61:1–25, 2010.

[3] Benjamin Bold. *Famous Problems of Geometry and How to Solve Them.* Dover, New York, 1982.

[4] James Robert Brown. *The Philosophy of Mathematics: A Contemporary Introduction to the World of Proofs and Pictures.* Routledge, London, 2nd edition, 2008.

[5] Florian Cajori. *A History of Mathematical Notation.* Dover, New York, 1993.

[6] Stephen Chrisomalis. *Numerical Notation: A Comparative History.* Cambridge University Press, Cambridge, 2010.

[7] Mark Colyvan. Mathematical recreation versus mathematical knowledge. In Mary Leng, Alexander Paseau, and Michael Potter, editors, *Mathematical Knowledge*, pages 109–122. Oxford University Press, Oxford, 2007.

[8] Mark Colyvan. There is no easy road to nominalism. *Mind*, 119:285–306, 2010.

[9] Mark Colyvan. The ins and outs of mathematical explanation. *Mathematical Intelligencer*, forthcoming.

[10] Helen De Cruz and Johan De Smedt. Mathematical symbols as epistemic actions. *Synthese*, 190:3–19, 2013.

[11] Catarina Dutilh Novaes. *Formal Languages in Logic: A Philosophical and Cognitive Analysis*. Cambridge University Press, Cambridge, 2012.

[12] Herbert B. Enderton. *Elements of Set Thenry*. Academic, New York, 1977.

[13] Mohan Ganesalingam. *The Language of Mathematics: A Linguistic and Philosophical Analysis*. Springer, Heidelberg, 2013.

[14] Timothy Gowers and Michael Neilson. Massively collaborative mathematics. *Nature*, 461:879–881, 2009.

[15] Philip Kitcher. Explanatory unification. *Philosophy of Science*, 48:507–531, 1981.

[16] John E. Littlewood. *A Mathematician's Miscellany*. Methuen, London, 1953.

[17] Paolo Mancosu and Johannes Hafner. Unification and explanation: A case study from real algebraic geometry. In Paolo Mancosu, editor, *The Philosophy of Mathematical Practice*, pages 151–178. Oxford University Press, Oxford, 2008.

[18] Mark Steiner. Mathematics, explanation, and scientific knowledge. *Noûs*, 12:17–28, 1978.

[19] Zach Weber. Figures, formulae, and functors. In Sun-Joo Shin and Amirouche Moktefi, editors, *Visual Reasoning with Diagrams*, Studies in Universal Logic, pages 153–170. Springer, Basel, 2013.

Received \jreceived