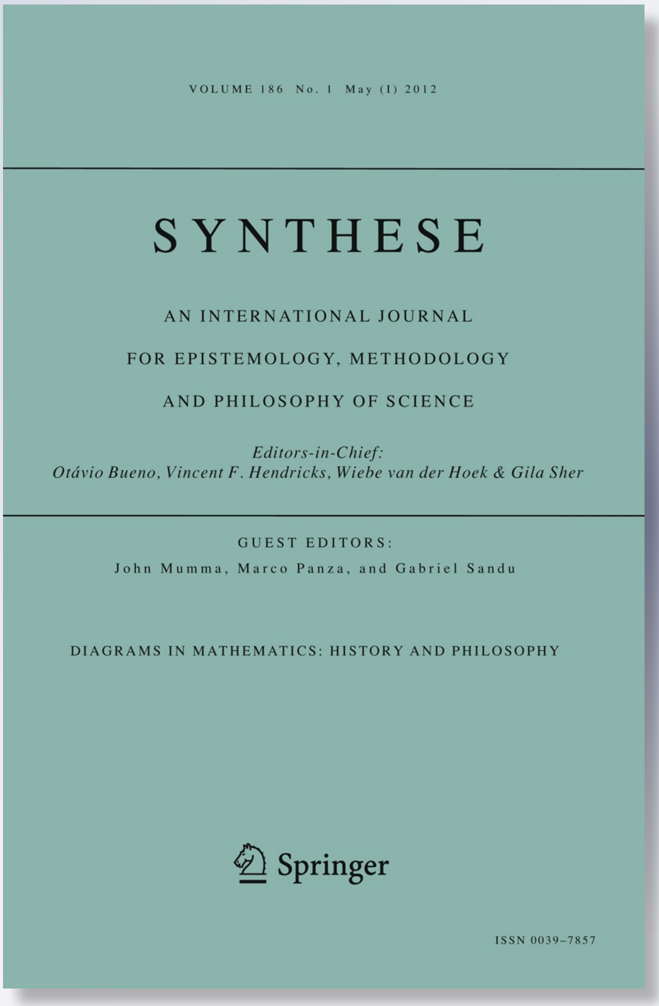
And so on . . . : reasoning with infinite diagrams

Solomon Feferman

Synthese
An International Journal for Epistemology, Methodology and Philosophy of Science

ISSN 0039-7857
Volume 186
Number 1

Synthese (2012) 186:371-386
DOI 10.1007/s11229-011-9985-6



Springer

Synthese (2012) 186:371–386
DOI 10.1007/s11229-011-9985-6

And so on ...: reasoning with infinite diagrams

Solomon Feferman

Received: 3 February 2010 / Accepted: 13 July 2010 / Published online: 3 August 2011

Abstract This paper presents examples of infinite diagrams (as well as infinite limits of finite diagrams) whose use is more or less essential for understanding and accepting various proofs in higher mathematics. The significance of these is discussed with respect to the thesis that every proof can be formalized, and a "pre" form of this thesis that every proof can be presented in everyday statements-only form.

Keywords Diagrammatic reasoning · Infinite diagrams · Formalizability thesis

1 Introduction

A proof of a theorem in mathematics is what we require to convince ourselves and others of the truth of the statement made by the theorem. Here 'truth' is taken in its *prima facie* sense, i.e., the notions involved in the statement of the theorem are supposed to be meaningful, and if it is to be truth *for us*, we are supposed to understand the meaning of those notions. In order to be convinced of a proof, one must follow the argument and check the various steps for ourselves, making use not only of what is given in the proof itself but what is required from background knowledge, i.e., previous statements that we have already accepted to be true on some ground or other. And that background knowledge may require understanding other notions not explicitly involved in the statement of the theorem. So both background knowledge

This paper is based in part on a lecture delivered to the Workshop on Diagrams in Mathematics, Paris, October 9, 2008. Slides for that talk and another one on the same topic for the Logic Seminar, Stanford, February 24, 2009 are available at http://math.stanford.edu/~feferman/papers/Infinite_Diagrams.pdf.
I wish to thank Jeremy Avigad, Michael Beeson, Hourya Benis, Philippe de Rouilhan, Wilfrid Hodges, and Natarajan Shankar as well as the two referees for their comments on a draft of this article. I wish also to thank Ulrik Buchholtz for preparing the LaTeX version of this article.

S. Feferman (✉)
Department of Mathematics, Stanford University, Stanford, CA 94305–2125, USA
e-mail: feferman@stanford.edu; s.feferman@gmail.com

and the understanding of meanings is an essential component of what it takes to accept a given proof. Even given that, it is possible to go through the steps of a given proof and not understand the proof itself. That is a different level of understanding, which, when successful, leads one to say, "Oh, I see!" In other words, this "really understanding the proof" is a special kind of insight into how and why the proof works. And that is necessary if one wants to follow proofs of related theorems and contribute to the subject by creating new proofs oneself. It follows that one cannot truly be a consumer and producer of mathematics without achieving real understanding of the arguments.[1]

Many proofs that mathematicians give rely on diagrams to a significant extent. They were ubiquitous in Greek geometry and in early analysis, but doubts were cast on their validity among other reasons because they might not be "typical": the worry was that they might in one way or another lead one to draw conclusions not justified by the hypotheses of the theorem to be demonstrated or be incomplete by not dealing with all possible cases. The process of rigorization of mathematics in the nineteenth century has supposedly led to their elimination in principle from modern mathematics. But the practice of reliance on diagrams in various ways is still integral to the presentation of mathematical proofs of all sorts, even outside of geometry and analysis. That is because such use of diagrams is part of what we make use of in arriving at real understanding of various proofs.

The concept of a mathematical diagram used here is a rather general one; it is supposed to be a representation of an abstract mathematical configuration on a two-dimensional surface consisting of points, lines, curves, arrows, with labels, marks, shaded areas, and so on. Among the "and so on" there may be special features such as the use of broken lines to represent three-dimensional or even higher-dimensional aspects of the given configuration, of dotted lines to represent a construction to be made at a certain point in the argument, of parallel lines viewed in perspective so that they meet at "infinity," etc. In view of all this it is genuinely questionable whether one can say precisely what constitutes a mathematical diagram, let alone a diagram in general.

Most theorems in mathematics state a fact about infinitely many objects of a certain kind, e.g., all triangles. But the diagram used in a proof only represents one such object, and as already mentioned it is an issue whether the particular representation taken is typical, i.e., does not have features which are not shared by all the objects under consideration. This is a frequent concern when dealing with proofs of geometric theorems that rely to a significant extent on diagrams. What I want to do here is bring attention to a different kind of diagram that is ubiquitous in modern mathematics, in which there is a *single infinite configuration* under consideration, and what is exhibited in the diagram is *a typical finite part of that configuration*, with the balance indicated by the use of ellipsis, i.e., dots '…' expressing 'and so on' or 'and so forth' in some way. The consideration of such infinite diagrams is interesting because they enlarge the question of what makes a diagram typical. Moreover, I shall argue that there are certain

[1] Wilfrid Hodges has remarked to me that he is not alone in having said, "Oh, I see!" to proofs that he later realized he hadn't understood. He went on to say that "[a]rguably one should define 'really understanding a proof' in terms of ability to paraphrase or adapt it, or apply it, or answer objections to it, and maybe other kinds of reactions to it. If this leads to no clear dividing line between really understanding and not really understanding, that seems to me to fit our experience."

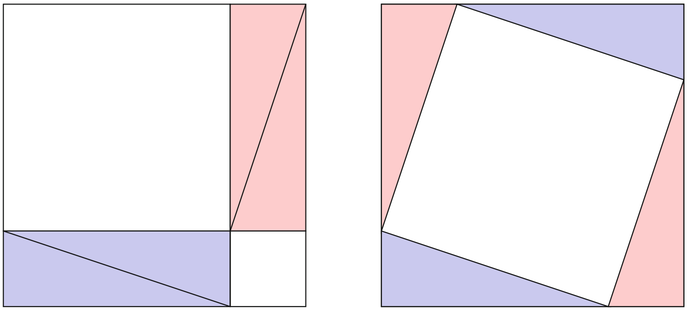


Fig. 1 A diagrammatic proof of Pythagoras' Theorem

proofs in modern mathematics where the use of such infinite diagrams is essential, i.e., it is not possible to even follow the proof without consulting the diagram at practically every step of the way. In fact, there are certain theorems whose statement can't even be understood without reference to such a diagram. This raises a *prima facie* problem for the thesis that every mathematical proof can be formalized. Actually, it already raises the problem for the thesis that every proof of a mathematical theorem that may involve diagrams and other possible devices can be replaced by an everyday statements-only proof. I shall call these the *Formalizability Thesis* (FT), and the *Pre-Formalizibility Thesis* (PFT), respectively. Actually, the problems raised for these theses are already issues for the essential use of finite diagrams for the understanding of various proofs, so we need to attend to the question of what further problems, if any, may be raised by the use of infinite diagrams in higher mathematics.

There has been a great resurgence of interest in reasoning with diagrams in mathematics (and visual reasoning more generally) especially in the last couple of decades, and the literature dedicated to that has become quite extensive. The survey article by Mancosu (2005) provides an informative entry to the subject. Inevitably, what I have to say in the following concerning the use of diagrams in general is no doubt already to be found in that literature or overlaps it to a considerable extent. But as far as I know, the considerations here concerning the modern uses of infinite diagrams are novel.

Before starting, one thing I want to emphasize about diagrams of all sorts, whether finite or infinite, is that we should not think of them in the way usually done, as static figures when reading a text, but rather as they might be presented in a lecture, constructed in stages.[2] In that very process our understanding of what is going on begins to feed into our understanding of the proof in which the diagram participates. And then when it is completely before us, we may retrace various of its aspects to further fill out that proof. Here's a classic example of that, with a proof, in stages, of Pythagoras' Theorem (one of the hundreds of known proofs). The figure is taken from Roger Nelsen's *Proofs Without Words* (1993).

One starts with a diagram of a right triangle, say the lightly shaded one in the lower right hand corner of the right hand diagram in Fig. 1, and then constructs the square on

[2] A closely related point concerning Euclidean diagrammatic demonstrations is made by Manders (2008a, 3.1.1, pp. 68–69).

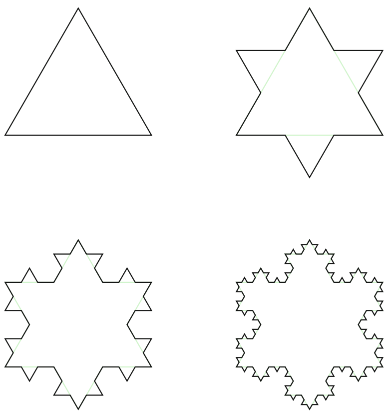
its hypotenuse. Following that, we add three more copies of the original triangle, each having the same square on the hypotenuse, to form a larger square whose side equals the sum of the two sides of the original triangle. Its area is the area of the square on the hypotenuse plus four times the area of the initial triangle. Then we form a different representation of those four triangles in that large square by reassembling them as shown in the left hand diagram of Fig. 1. This makes evident that the area of the large square is also equal to the sum of the squares on the sides of the initial triangle plus four times the square of its area. By subtraction from this equality we obtain Pythagoras' Theorem. When carrying out that demonstration in class on a blackboard, we hardly use all those words after the initial construction and reassembling, since the proof is almost evident by inspection. (Alternatively, we may work from left to right.)

The plan of the paper is as follows. Before going to infinite diagrams *per se*, in Sect. 2 I shall give examples of diagrams that are infinite limits of finite geometrical diagrams; in those cases it is difficult to visualize the limiting figure, but the process by which they are approximated is very clear. Section 3 presents three examples of the use of infinite diagrams in modern mathematics, one from set theory, one from model theory, and one from homological algebra. Finally, Sect. 4 discusses the relevance of these examples to the theses PFT and FT.

2 Finite diagrams with infinite limits

The first kind of infinite diagrams that I want to consider are those that are difficult to picture in and of themselves but are easily conceived as limits of finite diagrams obtained by iterating certain constructions. A simple example is provided by the familiar Koch "Snowflake" that is used to demonstrate the existence of a bounded continuous closed curve with no finite length and no tangent at any point. As shown in Fig. 2, it is the limiting curve of a sequence of polygons beginning with an equilateral triangle of side 1. The sequence is described inductively: at each stage, one simultaneously divides each side of the polygon before us into three equal segments, then builds an equilateral triangle on the middle segment, and finally deletes the base of the new

Fig. 2 Finite stages of the Koch "snowflake"



triangle except for its endpoints. Since the length of the circumference of this figure at each stage is multiplied by $4/3$, and since $(4/3)^n$ approaches infinity, the limiting curve has no finite length. The first four terms of this sequence are shown in Fig. 2, though three terms would have been sufficient to visualize where the process is heading.

The Koch snowflake is just one of a number of counter-intuitive or "pathological" functions and figures that emerged in the latter part of the nineteenth century in the process of the rigorization of analysis and the development of set theory and point-set topology. Of these Henri Poincaré wrote in 1906:

> Logic sometimes breeds monsters. For half a century there has been springing up a host of weird functions, which seem to strive to have as little resemblance as possible to honest functions that are of some use. No more continuity, or else continuity but no derivatives, etc. …Formerly, when a new function was invented, it was in view of some practical end. Today they are invented on purpose to show our ancestors' reasonings at fault, and we shall never get anything more out of them. (Poincaré 1952, p. 125)

Implicitly referred to here is the well known example due to Weierstrass of an analytically defined continuous but nowhere differentiable function.

In contrast to Poincaré, Hans Hahn argued against the dependence on intuition in mathematics in his famous essay, "The crisis in intuition" (Hahn 1933). He there pointed to a number of mathematical "monsters" to support his critique, such as a simplification due to Hilbert of Peano's space filling curve, an example due to Brouwer of a "map" of three "countries" which meet each other at every point of their boundaries, and a curve due to Sierpiński which intersects itself at every point. Another topological monster which could have been mentioned by Hahn is the so-called Alexander "horned" sphere which is homeomorphic to the unit sphere in three dimensions yet whose complement is not simply connected.

As I argued in Feferman (2000), the purpose of such monsters in the development of modern mathematics was to show that when one makes precise in analytic terms our intuitive notions of continuity, curve, tangent, boundary, connectedness, etc., ordinarily expected consequent properties don't necessarily hold. Thus if it is smooth curves about which one wants to obtain results, a hypothesis of differentiability must be added, and so on. Ironically, as with the Koch snowflake above, the understanding of how such counter-intuitive examples are generated makes use of intuitively clear finite diagrams or pictures which approach the "monster" in the limit. For example, the Peano-Hilbert example of a space-filling "curve" is the limit of curves that first go through every quadrant of the unit square, then modified to go through every sub-quadrant, and so on. The Sierpiński "curve" begins by deleting the interior of an inscribed equilateral triangle within an initial such triangle; the required figure is the skeleton of what's left in the limit of iterating this process. The Alexander horned sphere is formed by successively growing "horns" from the unit sphere that are almost interlocked and whose end points approach each other. This can be visualized by posing one's thumb and forefinger of each hand toward those of the other hand as if one is going to interlock them, then imagine growing a smaller thumb and forefinger on the end of each of these, and so on.

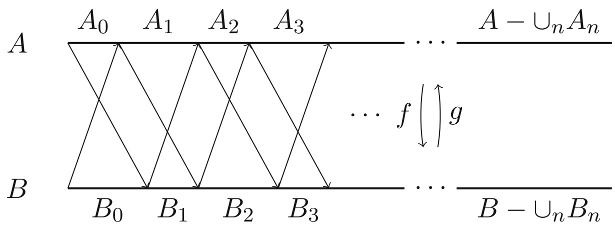


Fig. 3 A diagrammatic proof of the Cantor-Bernstein Theorem

The $1/3$ construction procedure in the Koch snowflake (made public in 1904) may have been suggested by the Cantor construction in the 1880s of an uncountable nowhere dense subset of the closed interval [0, 1] having Lebesgue measure 0 that is obtained by deleting successive middle open thirds. Later constructions, in 2 and 3 dimensions, respectively, of uncountable nowhere dense sets of measure 0 are the so-called Sierpiński carpet and the Menger sponge. The notion of measure was introduced in part to serve as a precise extension to more or less arbitrary sets of the intuitive notions of length, area and volume. All are examples of what are now called fractals, popularly enjoyed for their unusal semi-visual properties.

3 Proofs appealing to representations of infinite diagrams

Let us now turn to infinite diagrams which can be visualized in full, in contrast to those of the preceding section, though they may also involve the iteration of certain constructions. Three examples are given here, the first from set theory, the second from model theory, and the third from homological algebra.

We begin with a proof of the Cantor-Bernstein Theorem.[3] Given two sets A, B one defines $A \preceq B$ to hold if there is a one-one mapping of A into B, and $A \equiv B$ if there is a one-one mapping of A onto B. The Cantor-Bernstein Theorem tells us that if $A \preceq B$ and $B \preceq A$, then $A \equiv B$. In the diagram used for the proof in Fig. 3, A is represented by a broken line above and B by a broken line below, with the reason for the breaks explained in the process of the proof. One begins by taking f to be a $1-1$ map of A into B and g to be a $1-1$ map of B into A. To proceed, we look alternately at what each of f, g misses on the other side, beginning with $A_0 = A - g(B)$ and $B_0 = B - f(A)$. Then A_0 can be matched up on the B side with $B_1 = f(A_0)$, while B_0 can be matched up on the A side with $A_1 = g(B_0)$, so $A_0 \equiv B_1$ by f while $A_1 \equiv B_0$ by the inverse g^{-1} of g. Moving on, we define A_2 as $g(B_1)$ and B_2 as $f(A_1)$, and then A_3 as $g(B_2)$ and B_3 as $f(A_2)$. This leads us to define A_n and B_n in general for all n, in such a way that $A_{2n} \equiv B_{2n+1}$ by f and $A_{2n+1} \equiv B_{2n}$ by g^{-1}. Hence the union of the A_n's is in $1-1$ correspondence with the union of the B_n's. But those unions might not exhaust the sets A and B. That can hold for example, if both A_0 and B_0 are countable while A and B are uncountable; then the respective unions are countable while their complements are uncountable. The part to the right of the ellipsis on each side of the

[3] Also called the Schröder-Bernstein theorem. According to Kuratowski and Mostowski (1968, p. 190), its first correct proof was obtained by F. Bernstein and published in a book by E. Borel in 1898.

diagram represents that possibly non-empty complement. To finish off the proof, one argues that f is a $1-1$ map of A minus the union of the A_n's onto B minus the union of the B_n's, since no element of the latter can be caught as the image by f of some A_{2n} or by g^{-1} of some A_{2n+1}.[4]

The next example is drawn from the subject of what are called amalgamation theorems in model theory as exposited in Hodges (1997, pp. 134–149). I assume some basic knowledge of the subject as needed for the following notation: '$\mathcal{L}$', '$\mathcal{L}_1$', '$\mathcal{L}_2$', …are used for first order relational languages, and 'A', 'B', …with or without subscripts or superscripts are used for $\mathcal{L}$-structures for various $\mathcal{L}$. Given structures A_0, A_1 for the same language, we write $A_0 \equiv A_1$ when A_0 and A_1 are elementarily equivalent, and $A_0 \preceq A_1$ when A_0 is an elementary substructure of A_1. Then we write $A_0 \to A_1$ when there exists a substructure A_0' of A_1 for which $A_0 \equiv A_0'$ and $A_0' \preceq A_1$. A basic result that is used in the main result below is the Tarski-Vaught Theorem according to which if $A_0 \preceq A_1 \preceq A_2 \preceq \cdots$ and $A = \cup_n A_n$, then each $A_n \preceq A$.

Amalgamation theorems apply to the case when we have two languages $\mathcal{L}_1$ and $\mathcal{L}_2$ with a common non-empty sublanguage $\mathcal{L} = \mathcal{L}_1 \cap \mathcal{L}_2$ and structures A in $\mathcal{L}_1$ and B in $\mathcal{L}_2$ that are somehow to be related over $\mathcal{L}$. Throughout the following, the A structures with or without subscripts or superscripts are in $\mathcal{L}_1$ and similarly the B structures are in $\mathcal{L}_2$. We write '$A|\mathcal{L}$' and '$B|\mathcal{L}$' for the respective restrictions of A and B to $\mathcal{L}$. Then the preceding notation is extended so that '$A \equiv B$' is written to mean that $A|\mathcal{L} \equiv B|\mathcal{L}$, while '$A \preceq B$' is written to mean that $A|\mathcal{L} \preceq B|\mathcal{L}$, and '$A \to B$' is written to mean that $A|\mathcal{L} \to B|\mathcal{L}$. We assume proved the Elementary Amalgamation Theorem (Hodges 1997, p. 135, Theorem 5.3.1) by which if $A_0 \equiv B_0$, then there exists A_1 with $A_0 \preceq A_1$ and $B_0 \to A_1$. This is pictured in Fig. 4.

It follows immediately that under the same conditions there exist A_1 and B_1 such that $A_0 \preceq A_1$ and $B_0 \preceq B_1$ and $B_0 \to A_1$ and $A_1 \to B_1$. This is illustrated in Fig. 5.

We are now in a position to construct an infinite diagram to prove the Strong Amalgamation Theorem (due to Abraham Robinson and exposited in Hodges 1997, p. 147, Theorem 5.5.1): Given $A \equiv B$, there exists a structure C in $\mathcal{L}_1 \cup \mathcal{L}_2$ such that $A \to C$ and $B \to C$, i.e., $A \to C|\mathcal{L}_1$ and $B \to C|\mathcal{L}_2$. The proof appeals to Fig. 6.

In Fig. 6 we write 'A_0' for 'A' and 'B_0' for 'B,' and use the diagram of Fig. 4 to construct suitable A_1 and B_1 satisfying the pictured relations and then the same to construct suitable A_2 and B_2, and so on. Let $A_\omega = \cup_n A_n$ and $B_\omega = \cup_n B_n$. By the Tarski–Vaught Theorem, each $A_n \preceq A_\omega$ and each $B_n \preceq B_\omega$. Also $A_\omega|\mathcal{L} \equiv B_\omega|\mathcal{L}$ by the successive $\to$-relations between the A_n's and the B_n's. So we can construct a C in $\mathcal{L}_1 \cup \mathcal{L}_2$ that agrees with A_ω up to isomorphism on $\mathcal{L}_1$ and agrees with B_ω up to isomorphism on $\mathcal{L}_2$, as required. By the way, from the Strong Amalgamation Theorem one quickly infers Craig's Interpolation Theorem and Robinson's Consistency Theorem.

My final example involving infinite diagrams comes from homological algebra. Such diagrams are ubiquitous in that subject as they are in combinatorial topology;

[4] I don't remember where I first saw this use of an infinite diagram for the proof of the Cantor-Bernstein Theorem. Another diagram is used in the proof of a lemma for the theorem given in Hrbacek and Jech (1999), p. 67. Still another diagram is used in an automated proof of the theorem in Barker-Plummer et al. (1996); there the diagram is used to provide strategic information to the theorem prover employed (GROVER).

$$\begin{tikzcd} A_0 \arrow[rr, phantom, "\preccurlyeq"] & & A_1 \\ & B_0 \arrow[ur] & \end{tikzcd}$$

Fig. 4 Diagram for elementary amalgamation

$$\begin{tikzcd} A_0 \arrow[rr, phantom, "\preccurlyeq"] & & A_1 \arrow[dr] & \\ & B_0 \arrow[ur] \arrow[rr, phantom, "\preccurlyeq"] & & B_1 \end{tikzcd}$$

Fig. 5 One alternation of elementary amalgamation

$$\begin{tikzcd} A_0 \arrow[rr, phantom, "\preccurlyeq"] & & A_1 \arrow[dr] \arrow[rr, phantom, "\preccurlyeq"] & & A_2 \arrow[dr] \arrow[rr, phantom, "\preccurlyeq"] & & \cdots & \\ & B_0 \arrow[ur] \arrow[rr, phantom, "\preccurlyeq"] & & B_1 \arrow[ur] \arrow[rr, phantom, "\preccurlyeq"] & & B_2 \arrow[rr, phantom, "\preccurlyeq"] & & \cdots \end{tikzcd}$$

Fig. 6 Infinitely iterated alternation of elementary amalgamation

a classic reference is Mac Lane (1975). For a quick illustration, I here follow three pages from an introductory text, Jans (1964), pp. 27–29, devoted to an explanation and the first part of a proof appealing to a certain infinite diagram of what is called the Exact Sequence of Homology Theorem.[5] This deals with abstract complexes given by an infinite sequence of modules C_n over a ring R, where n is in $\mathbb{Z}$, the set of integers.[6] Such a complex is given by a collection of R homomorphisms, $d_n : C_n \to C_{n-1}$, called differentials, such that for each n, $d_n d_{n+1} = 0$. For any such sequence, we have $\mathrm{Ker}(d_n) \supseteq \mathrm{Im}(d_{n+1})$, and we can form the homology groups $H_n(C) = \mathrm{Ker}(d_n)/\,\mathrm{Im}(d_{n+1})$. Homological algebra is concerned with information about these groups for various complexes. One defines a complex map $f : A \to C$ between any two complexes, to be a collection of R homomorphisms $f_n : A_n \to C_n$ such that the following diagram is commutative

$$\begin{tikzcd} \cdots \arrow[r] & A_n \arrow[r] \arrow[d] & A_{n-1} \arrow[r] \arrow[d] & \cdots \\ \cdots \arrow[r] & C_n \arrow[r] & C_{n-1} \arrow[r] & \cdots \end{tikzcd}$$

where the vertical homomorphisms are given by the f_n's and the horizontal ones are given by the d_n's, i.e., for each n, $f_{n-1}d_n = d_n f_n$ (superscripts on maps attached to specific complexes are dropped when there is no ambiguity). It is then shown by an easy argument that whenever this holds, the map f induces (for each n) an R homomorphism, $f_* : H_n(A) \to H_n(C)$. By an exact sequence of complexes

$$0 \to A \to B \to C \to 0 \tag{1}$$

[5] I only discovered after choosing this example that Rav (2007), p. 298, footnote 12, had pointed to the very same result and essential use of a diagram.

[6] In applications, $C_n = 0$ for all $n < 0$.

$$
\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
\cdots \arrow[r] & A_{n+2} \arrow[r] \arrow[d] & A_{n+1} \arrow[r, "d"] \arrow[d] & A_n \arrow[r, "d"] \arrow[d, "j"] & A_{n-1} \arrow[r, "d"] \arrow[d] & A_{n-2} \arrow[r] \arrow[d, "j"] & \cdots \\
\cdots \arrow[r] & B_{n+2} \arrow[r] \arrow[d] & B_{n+1} \arrow[r, "d"] \arrow[d] & B_n \arrow[r] \arrow[d, "\pi"'] & B_{n-1} \arrow[r] \arrow[d] & B_{n-2} \arrow[r] \arrow[d] & \cdots \\
\cdots \arrow[r] & C_{n+2} \arrow[r, "d"] \arrow[d] & C_{n+1} \arrow[r, "d"] \arrow[d] & C_n \arrow[r] \arrow[d] \arrow[uur, dashed] & C_{n-1} \arrow[r] \arrow[d] & C_{n-2} \arrow[r] \arrow[d] & \cdots \\
 & 0 & 0 & 0 & 0 & 0 &
\end{tikzcd}
$$

Fig. 7 Connecting homomorphism for exact sequence of homology

where $j: A \to B$ and $\pi: B \to C$, is meant one where for each n, the sequence

$$0 \to A_n \to B_n \to C_n \to 0 \tag{2}$$

is exact, i.e., $\mathrm{Ker}(\pi_n) = \mathrm{Im}(j_n)$ for each n. The Exact Sequence of Homology Theorem states that whenever $0 \to A \to B \to C \to 0$ is an exact sequence then there is an induced "long" exact sequence

$$
\begin{tikzcd}
\cdots \arrow[r] & H_n(A) \arrow[r] & H_n(B) \arrow[r] & H_n(C) \arrow[dll, rounded corners] & \\
 & H_{n-1}(A) \arrow[r] & H_{n-1}(B) \arrow[r] & H_{n-1}(C) \arrow[r] & \cdots
\end{tikzcd}
$$

given by a sequence of homomorphisms $\theta: H_{n+1}(C) \to H_n(A)$ called the connecting maps. The proof occupies three pages in Jans (1964), 29–31, and appeals repeatedly to the diagram on p. 29, reproduced here as Fig. 7.

Of this, Jans says that the dashed arrow from C_n to A_{n-1} is used to indicate the path to follow for the construction of the connecting map θ. While that proof—which I shall not reproduce here—is written out fully in symbols, anyone who studies it can hardly deny that the diagram in Fig. 7 is absolutely indispensable for understanding how it proceeds by "diagram chasing," i.e., the demonstration that the composition of maps along various paths from a given node to another in the diagram always gives the same result.[7] This is completely typical of arguments in homological algebra, combinatorial topology and modern algebraic geometry.

Note that from a logical point of view there is an essential difference between the infinite diagrams in Figs. 3 and 6 and that in Fig. 7, namely the former ones are constructed inductively while the latter is not. Indeed, one could say that what is at issue in Fig. 7 is just showing how the connecting map is determined for a finite diagram confined to the indices $n+1, n$ and $n-1$ in Fig. 7. However, a little further on in the subject one is led to inductively generated infinite diagrams, namely after the introduction in Jans (1964) p. 33 of the notion of a projective resolution

[7] In standard terminology, this is the statement that the diagram is commutative.

$$\cdots \to P_n \to P_{n-1} \to \cdots \to P_1 \to P_0 \to 0; \tag{3}$$

see for example the proof of the Simultaneous Resolution Theorem (*ibid*., p. 39). However, it would take us too far afield to explain the notions involved and the diagrams in question.

4 Proofs from diagrams and the Formalizability Thesis

All of this touches on a larger issue, namely *the thesis that every proof can in principle be formalized*, i.e., has a counterpart formal derivation in some formal system. For present purposes I will refer to this as the *Formalizability Thesis* (FT).[8] It should be clearly distinguished from the *Formalist Thesis*, according to which mathematics has no content but merely consists in following formal rules. The Formalizability Thesis is usually considered with respect to the *statements-only proofs* that one meets in practice, i.e., which consist of a sequence of statements given in natural language augmented by various kinds of mathematical symbolism. By the *Pre-Formalizability Thesis* (PFT) I mean *the thesis that every proof of a mathematical theorem that may involve diagrams and other possible devices can be replaced by an everyday statements-only proof*. In Sect. 1 I argued that for something to be a proof *for us* it is not sufficient that we be able to check through it step by step but we must also understand it as a whole. Moreover, there are proofs that make use of diagrams in a way that contributes substantially to the understanding of the proof. In particular, in the preceding I offered three examples of such from modern mathematics employing infinite diagrams, with the claim for the last—from homological algebra—that reference to the diagram is essential for its proof; moreover, that is just one of a multitude of proofs of that character. These sorts of examples raise a *prima facie* challenge to the Pre-Formalizability Thesis and thence to the Formalizability Thesis. Even examples from more elementary mathematics can be produced which raise the same challenges and use only finite diagrams. Nevertheless, I shall make a case in the following both for PFT and then for FT. In addition I shall explain why it is also necessary to consider proofs employing infinite diagrams in doing so.

4.1 The case for PFT

I have not seen any general discussions in the literature of the Pre-Formalizability Thesis but that does not mean that a case does not have to be made for it. Moreover, the case for FT clearly requires acceptance of PFT. Let us look at several example areas that may constitute specific challenges to PFT.

First of all, consider the "proofs without words" in Nelsen (1993). The initial ones among these make use of dot-patterns and other kinds of patterns that illustrate arithmetical identities such as ones of the form $f(1) + f(2) + \cdots + f(n) = g(n)$ for

[8] Some of my colleagues such as Michael Beeson and Natarajan Shankar say that FT is now universally granted, and so a defence of it is in effect beating a dead horse. That that is not the case is evidenced by various well-known critics of FT cited, for example, in Rav (1999; 2007).

various f and suitable g. The remarkable thing about these is that the typical examples such as those for the sum of the first n positive integers or first n odd numbers with some small n lead to a very quick conviction as to the truth of the statements for arbitrary n. The basis of that conviction may be only partially reasoned and when made explicit, may not be sufficient to be converted to a statements-only proof. But massive experience with these shows that in all cases they can be established by an inductive argument which may be quite different from the one suggested by the diagram. PFT only requires that we be able to *replace* a proof making use of diagrams by another one that is diagram-free.[9]

Euclidean geometry provides one of the most extensive body of diagram based proofs in mathematics. For many of these proofs, reference to the diagram is apparently indispensable in order to follow and understand the proof; take, for example, Euclid's proof in I.47 of the Pythagorean Theorem, or even the simpler proof described in Sect. 1 above. Nevertheless, in the critical reexamination of Euclidean geometry in the late nineteenth century by Pasch, Hilbert and others, the Pre-Formalizability Thesis was in effect claimed to hold. As referenced by Mancosu (2005) pp. 14–15, this is quite explicit in the case of Pasch; in the case of Hilbert it is only explicit in various of his lectures on geometry and not in his *Grundlagen der Geometrie* of 1899. Still, these are statements in principle, not a demonstration that PFT holds for Euclidean geometry. That has only been established quite recently in the work of Avigad et al. (2009) via a formal system E that is faithful to Euclid's and Euclidean style proofs. The system E takes Manders' 2008b distinction between exact (or metrical) and co-exact (or topological) attributes as its point of departure, and builds on experience from the Ph.D. dissertation of Mumma (2006) which provides a formal system Eu in which diagrams are still bona fide objects. Some examples are given in Sect. 4.2 (pp. 734–738) of Avigad et al. (2009) of how Euclidean proofs can systematically be replaced by informal statements-only proofs that can then be formalized directly in E.

The process of rigorization of analysis that began in the nineteenth century and that was followed by the rigorization of topology in the twentieth century and the subsequent rewriting of mathematics Bourbaki style would seem to make the case for PFT in these areas without much further ado. Still, as we saw in Sect. 2, various unusual figures ("monsters") were produced in the process to serve as counterexamples to putative (pre-rigorous) theorems, and one must test PFT for those cases. I have not checked the literature to verify that each of the figures in question does indeed have a description in analytic terms, but in general they are obtained as the limit of an inductively generated sequence of finite diagrams, each of which can, in principle, be described analytically, though that might require a certain amount of work.

Let's look, finally, at the three examples of infinite diagrams described in Sect. 3. There is no difficulty for each of these in simply replacing each proof by one in which those diagrams are eliminated in favor of symbols for the entities and their

[9] Jamnik (2001) deals with dot pattern proofs of such arithmetical identities in a systematic way. The idea is that one is led directly from small typical diagrams to formulate a conjecture leading to a program P that constructs for each n a formal proof $P(n)$ of the given identity. Then the general statement of the identity follows by the recursive ω-rule. However, each such P must still be supplemented by a proof of correctness at the meta level to arrive at a fully formal proof and that inevitably involves an inductive argument.

relationships that are partially represented in the diagrams, and the proofs may then be carried out in the everyday statements-only form. As we shall see, what is difficult in these cases, as compared to the ones above from more elementary mathematics, is not the verification of the Pre-Formalizability Thesis but rather the Formalizability Thesis.

Note that it is *not* claimed in any of the above cases that understanding and conviction are retained in the process of eliminating appeal to diagrams in favor of proofs in statements only form. The Pre-Formalizability Thesis makes no demands in that respect. Moreover, in practice students of such proofs often supply their own pictures and diagrams with which to gain the requisite insight and conviction.

4.2 The case for FT

It is the Formalizability Thesis rather than the Pre-Formalizability Thesis that has been the subject of extensive discussion and controversy. Historically, that had its origins in Frege's *Begriffsschrift* and was considerably bolstered by the work of Whitehead and Russell in the *Principia Mathematica*. Then at a more general level the idea of a formal system came to the forefront via Hilbert through his finitist consistency program, and for that reason the view is called by some *Hilbert's Thesis*; however, it should be understood independently of Hilbert's program.[10] Just what the thesis means without begging the question as to what a "proof" is is hard to say, but the idea is common enough, and has both many defenders and opponents. Some recent strong critiques of it which also form a guide to the relevant literature have been mounted by Rav (1999, 2007) and Pelc (2009), while—in a response to Rav (1999)—Azzouni (2004) argued for a version of formalism, according to which "[o]rdinary mathematical proofs indicate (one or another) mechanically checkable derivation of theorem from the assumptions those…proofs presuppose." (*ibid.*, p. 205).[11]

While there are very general precise explanations of what constitutes a formal system, a real difficulty in any defense of FT lies in saying just what formal system is to be associated with a given informal proof, i.e., what is to be taken for its language, axioms and rules of inference, and what it means to formalize a given proof in such a system.[12] Nevertheless there is an extensive body of experience in modern mathematical logic that can be appealed to, to flesh this out in a great variety of cases.

[10] In addition, Hilbert is often mistakenly referred to as a formalist.

[11] These particular discussions use 'proof' for the informal arguments found in mathematical practice and 'derivation' for their presumed counterparts in formal systems, but that terminology is by no means universal. For a careful discussion of formal vs. informal provability, see Leitgeb (2009).

[12] Philippe de Rouilhan observed that Frege and Russell each proposed a strong form of FT, namely that there is a single formal system in whose language every mathematical notion can be expressed and in which every mathematical theorem can be derived. We know the fates of their specific proposals due to inconsistency and incompleteness, respectively. In general, Gödel's incompleteness theorem undermines any proposal for such a strong form of FT. However, some claim that all mathematical notions can be defined in the language of set theory while others claim that they can all be defined in the language of category theory; just what such claims come to deserves further analysis. (In neither case, of course, is it claimed that some specific axiomatic system of set theory, resp. category theory, is sufficient to derive all of mathematics.)

One might argue that—granting PFT—the thesis has independent plausibility for informal proofs consisting solely of a reasoned sequence of statements that only involve words and symbols, since there is hardly any dispute as to the formalizability of individual statements in a suitable language. However, the difficulty lies in the steps from one statement to the next whose justification may be evident to the human mathematician specializing in the subject matter of the proof but that require extensive filling in, in order to create a fully formal derivation. And it is in this respect where the kind of reasoning behind the examples in Sect. 3 from modern infinitistic mathematics raises particular difficulties, because it is not in general simply logical microsteps that have to be inserted. Rather, in practice, the expert human mathematician routinely calls on a repertoire of prior notions, methods and results from his memory to readily recognize the validity of the steps in question. Depending on the mathematics of the proof in question, those notions, methods and results may be about sets and functions, or models and satisfaction, or modules and groups and homomorphisms, and so on. But they may also involve mathematics not explicitly present in the steps being filled in. For example, in the case of infinite diagrams, there is constant appeal to the indexing, and hence to background knowledge about the integers, including the use of inductive arguments. And supplying the detailed intermediate steps in a suitable formal system is by no means routine. However, it is here where the considerable experience in recent years with the mechanical verification of proofs comes in to give additional substance to the thesis. See for example, http://www.mizar.org/ for the Mizar system of proof checking in a formal system of set theory, and Nipkow et al. (2002) for the Isabelle higher-order logic proof-assistant. Note the use of the word 'assistant,' for the preparation of an informal statements-only proof for formal checking requires detailed guidance by the mathematician(s) in charge.[13]

Despite this kind of evidence, one must still give attention to the critical side and to the more specific question whether the kind of use of diagrams illustrated in this paper provides specific material for arguments *contra* the Formalizability Thesis. Rav's general criticism is of "the belief that through complete formalization (in a suitable formal language) mathematical proofs attain the optimum of certainty and reliability" (Rav 2007, p. 291). He points out that that view is not to be confounded with standards of rigor that, historically, have changed and evolved and varied from subject to subject within mathematics. By contrast, Rav identifies the true function of proofs within mathematical practice to lie in their interconnected role in the development of individual subjects. As he wrote in his earlier article, "Why do we prove theorems?",

> Proofs are for the mathematician what experimental procedures are for the experimental scientist: in studying them one learns of new ideas, new concepts, new strategies—devices which can be assimilated for one's own research and be further developed. (Rav 1999, p. 20)

The article Rav (2007) continues and elaborates these themes as part of his critique of Azzouni (2004). Pelc, on the other hand, is focused on the much more restrictive

[13] An interesting recent example is provided by the mechanical proof check in Avigad et al. (2007) of the Erdős-Selberg "elementary" proof of the prime number theorem also using Isabelle.

question, “Why do we believe theorems?”, as he entitles his 2009 paper. Without dismissing Rav’s points, he says that

> [n]evertheless, the role of proofs as a means of convincing the mathematical community of the validity of theorems is very important. While proofs can also serve other purposes, only proofs can directly serve this purpose. …[here] we are only interested in the ‘convincing’ role of proofs.” (Pelc 2009, p. 85)

And in this respect Pelc’s main criticism of the Formalizability Thesis is found in the abstract to his paper:

> The formalist [*sic*!] point of view maintains that formal derivations underlying proofs, although usually not carried out in practice, contribute to the confidence in mathematical theorems. Opposing this opinion, the main claim of the present paper is that such a gain of confidence obtained from any link between proofs and formal derivations is even in principle impossible in the present state of knowledge (Pelc 2009, p. 84).

For his argument, Pelc defines a natural number M that is so large that no theorem T whose shortest possible derivation in ZFC is of length greater than M will be mechanically checkable by a physically realizable process within anything like feasible time. And then he goes on to suggest that the proof by Wiles of Fermat’s Last Theorem could be a candidate for such T “in the present state of knowledge.” (Considering the great progress being made in the actual mechanical checking of proofs referred to above, this is rather incautious speculation.)

It is seen that the arguments of Rav and Pelc are not arguments against the Formalizability Thesis *per se*, but rather arguments to the effect that informal proofs serve a number of purposes that cannot be served by any supposed formalizations of them, first and foremost their role in convincing us of the truth of the statements which they purport to establish. In the introduction, I, too, took that to be the primary purpose of proofs and added that understanding proofs is a necessary part of that. In particular, in Sect. 2 and 3 I brought forth some examples of diagrams which play to some extent or other an essential role in gaining that understanding; indeed, I claimed that that is completely the case in the final example considered in Sect. 3 (and is so also for similar proofs throughout homological algebra and topology).[14] Nevertheless, I do not see that as an argument against the Formalizability Thesis.

In other words, I believe that the Formalizability Thesis should be given a very strict reading, namely that (i) every good proof has an underlying logical structure, (ii) that structure is completely analyzed in the derivation that formalizes the proof, and, finally (iii) that derivation assures the correctness of the theorem proved on the basis of the background assumptions expressed by the axioms and rules of the system in which

[14] Note that I have only been concerned here with the use of diagrams in more or less sophisticated mathematical reasoning. Besides the work of Avigad et al. (2009) on Euclidean geometry mentioned above there is also an extensive literature on systematic reasoning with diagrams outside of these areas; see, among others, Allwein and Barwise (1996); Jamnik (2001) and Shin and Lemon (2008) for an introduction to that. Other directions of work, such as Barwise and Etchemendy (1996) involve heterogeneous systems of reasoning, e.g., employing manipulations of icons on a computer screen.

the proof is formalized.[15] This is an *in principle* thesis that has nothing to do with conviction, understanding, or feasibility and it seems to me to be perfectly consistent with the view of the central and methodological virtues of proofs emphasized by the critics.[16]

The Formalizability Thesis is especially important for logicians since the claims for formalizability of various particular informal arguments are ubiquitous in our work. This goes back to Gödel's proof of his incompleteness theorems, especially the second theorem, on unprovability of consistency, the standard argument for which requires formalizing the proof of the first incompleteness theorem. And just as for Gödel's theorems, the significance of all the subsequent work in metamathematics for the potentialities and limitations of mathematical thought depends essentially on the extent to which the various formal systems of algebra, number theory, analysis and set theory that have emerged in that work account for extensive swaths of mathematical practice. The underlying logical structure of mathematics is an essential part of what makes it such a distinctive body of thought and it is exactly the Formalizability Thesis that allows us to say in precise terms a good deal—but by no means all—of what we are up to when we are doing mathematics.

References

Allwein, G., & Barwise, J. (Eds.). (1996). *Logical reasoning with diagrams*. New York: Oxford University Press.

Avigad, J., Dean, E., & Mumma, J. (2009). A formal system for Euclid's *elements*. *Review of Symbolic Logic, 2*(4), 700–768.

Avigad, J., Donnelly, K., Gray, D., & Raff, P. (2007). A formally verified proof of the prime number theorem. *ACM Transactions on Computational Logic 9*(1), Art. 2.

Azzouni, J. (2004). The derivation-indicator view of mathematical practice. *Philosophia Mathematica, 12*(2), 81–105.

Barker-Plummer, D., Bailin, S. C., & Ehrlichman, S. M. (1996). Diagrams and mathematics. In B. Selman & H. Kautz (Eds.), *Proceedings of the Fourth International Symposium on Artificial Intelligence and Mathematics* (pp. 14–17). Fort Lauderdale, FL.

Barwise, J., & Etchemendy, J. (1996). Heterogeneous logic. In G. Allwein & J. Barwise (Eds.), *Logical reasoning with diagrams* (pp. 179–200). New York, NY: Oxford University Press.

Feferman, S. (1979). What does logic have to tell us about mathematical proofs?. *The Mathematical Intelligencer 2*, 20–24. (Reprinted as Ch. 9 in Feferman 1998).

Feferman, S. (1998). *In the Light of Logic*. New York, NY: Oxford University Press.

Feferman, S. (2000). Mathematical intuition versus mathematical monsters. *Synthese, 125*(3), 317–332.

Hahn, H. (1933). The crisis in intuition. In *Empiricism, logic, and mathematics: philosophical papers, Vol. 13 of Vienna circle collection* (pp. 73–102). Dordrecht: D. Reidel Publishing Company

Hodges, W. (1997). *A shorter model theory*. Cambridge: Cambridge University Press.

[15] Note the use of the word 'good' in (i). Wilfrid Hodges has pointed out to me that there are examples of proofs that are basically faulty and that lead to faulty formalizations.

[16] Rav (2007) p. 309 approvingly quotes my paper, Feferman (1979) p. 22, where I wondered whether "a conversion [into mechanically checkable form] of really difficult and subtle proofs [is] possible without the human agent understanding in all details what is to be converted. And if he does understand 'in all details' isn't the battle over (since complete understanding subsumes checking)?" I admit that the view expressed was part of a general skepticism I had about the value of mechanical proof checking, but since then I have become more sanguine about the kind of work mentioned above. In any case, the quote itself is consistent with the strict view of the Formalizability Thesis proposed here.

Hrbacek, K., & Jech, T. (1999). *Introduction to set theory, Vol. 220 of monographs and textbooks in pure and applied mathematics* (3rd ed.). New York: Marcel Dekker Inc.,.

Jamnik M. (2001). *Mathematical reasoning with diagrams*, Vol. 127 of *CSLI Lecture Notes*. Stanford, CA: CSLI Publications. (With a foreword by J. A. Robinson).

Jans, J. P. (1964). *Rings and homology*. New York: Holt, Rinehart and Winston.

Kuratowski, K., & Mostowski, A. (1968). *Set theory.* (Translated from the Polish by M. Maczyński). Warsaw: PWN-Polish Scientific Publishers.

Leitgeb, H. (2009). On formal and informal provability. In O. Bueno & Ø. Linnebo (Eds.), *New Ways in the Philosophy of Mathematics.* (pp. 263–299). New York, NY: Palgrave Macmillan.

Mac Lane, S. (1975). *Homology*. Berlin: Springer-Verlag.

Mancosu, P. (2005). Visualization in logic and mathematics. In P. Mancosu, K. Jørgensen, & S. Pedersen (Eds.), *Visualization, explanation and reasoning styles in mathematics, Vol. 327 of Synthese Library* (pp. 13–30). Netherlands: Springer.

Manders, K. (2008). Diagram-based geometric practice. In P. Mancosu (Ed.), *The philosophy of mathematical practice* (pp. 65–79). New York, NY: Oxford University Press.

Manders, K. (2008). The Euclidean Diagram (1995). In P. Mancosu (Ed.), *The philosophy of mathematical practice.* (pp. 80–133). Oxford: Oxford University Press.

Mumma, J. (2006). *Intuition formalized: Ancient and modern methods of proof in elementary geometry.* Ph.D. thesis, Carnegie-Mellon University.

Nelsen, R. B. (1993). *Proofs without words: Exercises in visual thinking*. Washington, DC: Mathematical Association of America.

Nipkow, T., Paulson, L. C., & Wenzel, M. (2002) *Isabelle/HOL—A proof assistant for higher-order logic, Vol. 2283 of LNCS*. Berlin, New York: Springer.

Pelc, A. (2009).Why do we believe theorems? *Philosophia Mathematica, 17*(1), 84–94.

Poincaré, H. (1952). *Science and method* (Translated by Francis Maitland. With a preface by Bertrand Russell). New York: Dover Publications Inc.

Rav, Y. (1999).Why do we prove theorems? *Philosophia Mathematica, 7*(1), 5–41.

Rav, Y. (2007).A critique of a formalist-mechanist version of the justification of arguments in mathematicians' proof practices. *Philosophia Mathematica, 15*(3), 291–320.

Shin, S.-J., & Lemon, O. (2008). Diagrams. In E.N. Zalta (Ed.) *The Stanford encyclopedia of philosophy*. CSLI, winter 2008 edition. URL: http://plato.stanford.edu/archives/win2008/entries/diagrams/.