Who's Afraid of Mathematical Diagrams?

Silvia De Toffoli

Princeton University

1. Introduction

Diagrams have been in the mathematician's toolbox since antiquity. In ancient Greece, diagrams reflected the mathematics of the time, which dealt mostly with magnitudes, geometric figures, and proportions.[1] The wide variety of diagrams available to us now reflects our own mathematical concerns. Diagrams aid mathematicians in representing not only geometric figures but also topological objects and structural relations. Figure 1 shows a familiar *geometrical* diagram, a *topological* diagram, and an *algebraic* diagram.[2]
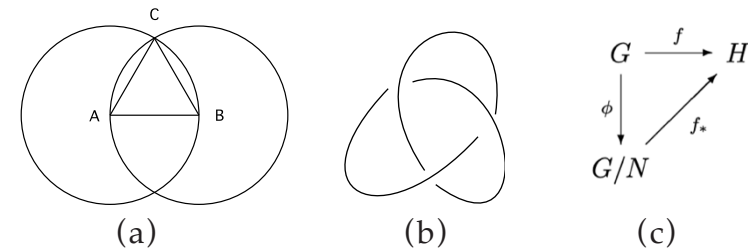

(a) (b) (c)

Figure 1. Three diagrams: (a) geometrical, (b) topological, and (c) algebraic

Despite the fact that they are common tools of mathematicians, diagrams have been understudied by contemporary mainstream philosophy of mathematics — which focuses mostly on foundational issues and on very general metaphysical and epistemological questions. Typically, mathematical diagrams are conceived of as, at best, superfluous illustrations and, at worse, dangerously idiosyncratic representations — ones that are helpful at times in amplifying our understanding but also capable of leading us astray. In an important recent book, John Burgess gives voice to this type of suspicion:

1. See, for example, Netz (1998; 1999) and Ferreirós (2016, Ch. 5).

2. Respectively, they are the diagram associated with Proposition I,1 of Euclid's *Elements* (which shows how to construct an equilateral triangle on a given segment), a diagram of the trefoil knot (the simplest non-trivial knot), and a commutative diagram expressing a relation between groups.

Some [diagrams] are surely no more essential to the proofs in whose midst they appear than the illustrations that accompanied many Victorian novels on first publication were essential to the literary value of those works. Other diagrams may play a more important role, especially perhaps in those types of abstract algebra where what is called "diagram chasing" is ubiquitous; though even here, the kinds of diagrams that are chased seem only to abbreviate information that could be put, more cumbersomely and in a way less easy to take in, in words. (2015, 98)

My aim in this paper is to address this form of skepticism about diagrams and to do justice to their use in contemporary mathematical practice. While I concede that *some* visual representations in mathematics are used as mere illustrations (even if they often play a significant heuristic role and thus are significantly different from the illustrations figuring in early Victorian novels), I nonetheless contend that this characterization hardly exhausts all such forms of visual representation in mathematics. In other words, while conceding that some diagrams are redundant devices that enhance understanding in pedagogic and research contexts alike, I argue that

> A) other diagrams do not serve an illustrative function and, on the contrary, form mathematical notational systems and therefore play a non-redundant role in proofs; and

> B) there is a plausible conception of proof according to which diagrams are not just abbreviations for more cumbersome non-diagrammatic displays but are in fact essential to those proofs.

In order to make the case for these claims, I discuss examples from two different domains in which diagrams are frequently used in proofs: low-dimensional topology and homological algebra.

Donald Davidson once said that a picture is "not worth a thousand words, or any other number. Words are the wrong currency to exchange for a picture" (1978, 47). I show that words are also the wrong currency to exchange for a diagram. In order to appreciate the effectiveness of diagrams, it is not enough to consider their informational content and how such content could be put into words. We have to consider how the articulation of that content (i.e., the organization of the content in the constitutive features) matters in practice. This can be done, for instance, by evaluating how it facilitates extracting information, carrying out specific inferences, and performing calculations — all functions that mathematicians generally recognize and accept.

In Sections 2 and 3, I argue that practitioners sometimes *think in and with diagrams* in ways that reveal those diagrams to be essential.[3] In Section 2, I focus specifically on diagrams that are not illustrations. I spell out specific conditions that diagrams must meet to count as elements of genuine mathematical notational systems. In Section 3, I consider diagrams that are not abbreviations of longer non-diagrammatic expressions. I show that there are plausible criteria of identity for proofs such that diagrams can play an essential role in proofs. An analogous debate in the realm of poetry concerns the so-called "heresy of paraphrase," the thesis according to which a poem cannot be paraphrased (or translated); "Poetry is what gets lost in translation," wrote Robert Frost. Something similar applies to diagrams. Like poems, diagrams can be transformed into linguistic displays. And, like poems, they cannot be translated. In a phrase: *inter-transformability does not imply inter-translatability*. In Section 4, I summarize my results.

2. Diagrams That Are Not Illustrations

The use of diagrams is one of the (previously neglected) topics in the philosophy of mathematics that has recently begun to receive

---

3. Elisabeth Camp (2007) makes a similar claim with respect to maps.

sustained attention.[4] The literature is now rife with detailed case studies of the uses of diagrams both in contemporary mathematics and in various historical periods — from Euclidean geometry to ancient China, knot theory, and the study of C*-algebras[5] — but tend not to address epistemological issues from a general perspective.[6] That is my goal here.

It is uncontroversial that diagrams can serve as powerful heuristic tools — tools that help us discover new propositions and amplify our understanding. What remains controversial, however, is whether diagrams can play a distinctly *justificatory* role. Indeed, since the end of the nineteenth century, diagrams have been regarded with outright suspicion by mathematicians and philosophers alike: the worry is that certain diagrams and visualizations can lead us astray. Though there are some cases in which this accusation is justified — for example, in encouraging the idea that we did not have to bother proving the Jordan curve theorem[7] or when geometric diagrams are not used rigorously and lead to overgeneralizations — in other cases it does not apply. In what follows, I use two examples to show that diagrams can contribute to the justificatory force of a proof.

2.1 Topological Diagrams

Let's start by considering diagrams of topological surfaces (i.e., two-dimensional manifolds). The *Möbius band* is the simplest non-orientable surface. It is usually represented either as a sketch of an object in space (Figure 2(a)) or as an *arrow diagram* (Figure 2(b); ignore the red dashed line for now). In Figure 2(b), the two vertical arrows have to be interpreted as *gluing instructions*: we have to imagine stretching the

square and twisting it in space to align the arrows and identify them. Imagining this process (or implementing it in a physical model made of paper), we obtain something like Figure 2(a) — the arrows can then be erased since they are now internal to the surface.
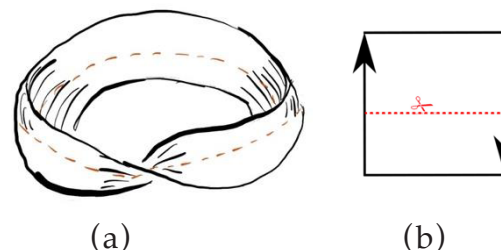

(a)   (b)

Figure 2. The Möbius band

Consider the following question: Which surface do we get if we cut the Möbius band along its core (indicated in the two representations by the dashed red lines)? Likely, Figure 2(a) will not help since our ability to imagine transformations in three dimensions is typically very limited and not especially reliable. We could use a paper model of the Möbius band and see what happens when we cut along its core with a pair of scissors. Alternatively, we can imagine *cutting and pasting* the arrow diagrams, creating the diagrammatic argument[8] in Figure 3.

> PROPOSITION: Cutting the Möbius band along its core
> gives rise to a cylinder.

---

4. See Mancosu (2008), Ferreirós (2016), and Carter (2019b).

5. See, for example, Manders (2008), Chemla (2018), De Toffoli and Giardino (2014), and Carter (2018), respectively.

6. Notable exceptions are Giaquinto (2007) and Carter (2019a).

7. This and similar points are mentioned in Hahn (1980, 93). The Jordan curve theorem might seem to be an obvious result, but it actually turns out to require sophisticated mathematical machinery in order to be proven.

8. With *diagrammatic argument*, I mean an argument in which diagrams play a prominent role. In our case, the argument consists in manipulating diagrams according to specific rules. A famous diagrammatic argument from ancient times is the one deployed by Socrates in Plato's *Meno* (81e–86c). The problem there is to construct a square with double the area of a given square – this can be done by a sequence of steps in which we manipulate a diagram in a way that is not dissimilar to the cutting and pasting method presented here; for an in-depth discussion of this example, see Giaquinto (2007, Ch. 4). Thanks to one of the anonymous referees for this analogy.
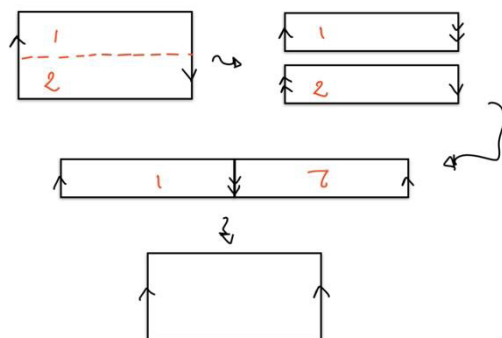
PROOF

Figure 3. Cutting a Möbius band along its core

The first step consists of cutting along the red dashed line. In order to conserve the original gluing instructions, we introduce the *double arrow* convention. In this way, we indicate which side of the first piece must be attached to which side of the second and in which direction. The second step consists of gluing the double arrows. To align the arrows, we flip the second piece. The third step is not a transition but a simplification. It consists of erasing the previously used gluing instructions and reshaping the rectangle. The final diagram is a rectangle with two sides identified in the same direction. It is a cylinder.

QED

This is a simple proof of a simple result. But there is a caveat. It is a proof only if the diagrams and diagrammatic transitions are interpreted correctly.[9] Cutting and pasting is not (only) an intuitive terminology to guide us in imagining diagrammatic transformations. These

9. See De Toffoli (2022) for a general analysis of when this is the case.

two operations can be cashed out in precise topological terms: gluing corresponds to forming a quotient space, and cutting can be thought of as the inverse operation of gluing.[10] The notation formed by arrow diagrams is a perspicuous diagrammatic notation that lets us carry out mathematical operations in an intuitive way. Moreover, it tracks relevant topological properties. For example, from Figure 2(b) we see that any band with an odd number of half-twists is also a Möbius band (Figure 4).
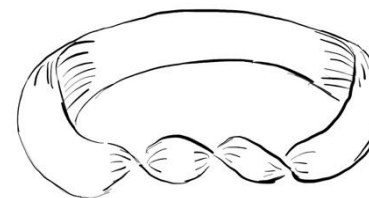

Figure 4. Möbius band with three half-twists

The number of twists is not an intrinsic property of the Möbius band, but one that derives from the way it is embedded in space. Abstracting from the particular embedding is useful in certain circumstances, but it hides how embeddings change with the various transformations. For example, the previous argument does not show that cutting the standard Möbius band along its core yields a *twisted* cylinder.

I called the argument in Figure 3 a "proof." It could be objected, however, that it is not a genuine proof but merely an exercise in applying transformations on a specific diagram. Fair enough. As you can see in the following theorem, arrow diagrams generalize and can be used to prove substantive results:

10. Roughly, quotient spaces are built by identifying groups of points of given spaces — in the case under examination, the sides with corresponding arrows are identified.

THEOREM 1. *(Classification of Surfaces.)* Any closed connected surface results from gluing the sides of a polygon according to one of the two gluing instructions in Figure 5 — which are represented by a label (and a color).[11] The pattern on the left corresponds to orientable surfaces, while the one on the right to non-orientable surfaces.
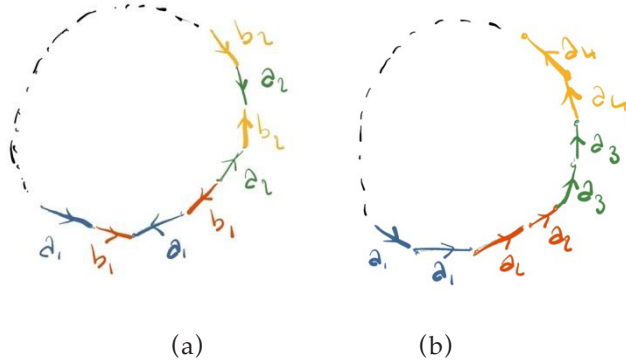

(a)   (b)

Figure 5. Classification of closed connected surfaces

ONE STEP OF THE PROOF

Figure 6 shows one step of a proof that appears in a graduate textbook in algebraic topology (Massey 1991, 23). The zigzagging convention introduces generality. It indicates that any combination of arrows could be positioned in its place.[12] The general strategy of the proof consists in showing that any collection of arrow diagrams representing a closed connected surface can be transformed into one of the two configurations in Figure 5.[13] The step in

---

11. A result needed for proving this theorem is that compact surfaces can be triangulated. For a proof, see Radó (1925).

12. Another diagrammatic proof is John Conway's Zero Irrelevancy Proof or "ZIP proof" (Francis and Weeks 1999), in which similar diagrams are cut and glued with *zippers*.

13. The proof thus assumes that we can start with some arrow diagrams. This is

Figure 6 is representative of the technique used. It starts with a diagram containing two pairs of alternating arrows. It proves that such a diagram can be transformed (without altering the surface it represents) into one with two adjacent pairs of arrows following the pattern in Figure 6(4) — which is a fragment of the pattern in Figure 5(a). Let's see how.
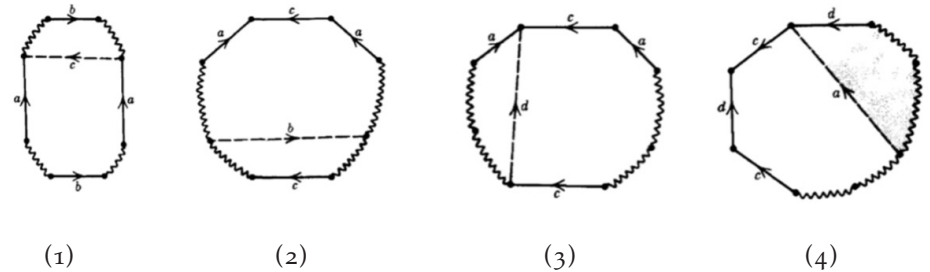

(1)   (2)   (3)   (4)

Figure 6. A step in the proof of the classification theorem for surfaces (Massey 1991, 23)

In order to go from (1) to (2), we cut along *c* and glue along *b*. The next two diagrams are equivalent. Figure 6(2) shows that *b* is now internal to the diagram. Figure 6(3) is obtained by erasing *b* and introducing the internal arrow *d* as a new cutting instruction. Figure 6(4) results from cutting along *d* and gluing along *a*.

QED

2.2 Syntax and Semantics of Arrow Diagrams

Let us pause briefly before tackling the second mathematical example. I want to draw out the main differences between the two representations for surfaces introduced above: arrow diagrams and sketches. The latter are evocative, pictorial representations of three-dimensional objects. Arrow diagrams are instead topological diagrams with precise

---

equivalent to saying that surfaces can be triangulated (Massey 1991, 14).

mathematical conventions. Diagrams have two central characteristics that pictorial sketches do not have: (i) they satisfy well-formedness conditions, and (ii) they are subject to precise rules of manipulation.

Let us first look at the syntax of arrow diagrams. I limit the discussion to closed connected surfaces. I will thus consider arrow diagrams formed by single polygons with all edges paired — the results can be easily generalized to non-connected surfaces. The primitive objects are *digons* (i.e., polygons with just two sides) with oriented arrows. As illustrated in Figure 7, there are just two possible basic diagrams, B_1 and B_2. As a matter of fact, the arrows' direction is not relevant per se; what counts is only the relative orientation of the arrows. This means that inverting the directions of both arrows of B_1 or B_2 would leave them unchanged. We can think of the orientation of each edge with respect to an orientation of the boundary circle. B_1 can then be characterized as the digon in which the two arrows have opposite directions (if you imagine tracing the boundary circle from a node, you are going to have to trace one of the arrows in reverse) and B_2 as the digon in which the two arrows have the same direction.
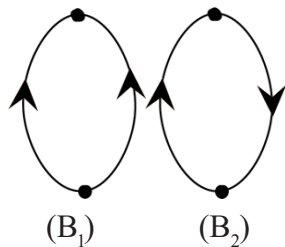

(B_1) (B_2)

Figure 7. Digons

All other arrow diagrams can be obtained from these two basic diagrams, B_1 and B_2, by adding couples of oriented edges. Again, when adding two edges we will have just two choices of orientation: they will go either in the same direction or in opposite directions along the

boundary circle. The set of well-formed arrow diagrams is the smallest set such that:

    1) B_1 and B_2 are in D.

    2) If the arrow diagram P is in D, and the arrow diagram P′ is obtained from P by adding two matching oriented edges, then P′ is in D.

In order to obtain P′ from P, we add two matching oriented edges, either with the same or with opposite orientation. We can position each of them in the place of a vertex. They can, therefore, be positioned one after the other or separated by one or more preexisting edges.

Figure 8 shows how we can add a pair of oriented edges to B_1, the first digon in Figure 7. There are four possibilities, corresponding to the choices of (i) whether the arrows are in the same or different direction and (ii) whether we insert the arrows one after the other or in different nodes — since the nodes of the digon are equivalent, there are just two possibilities.
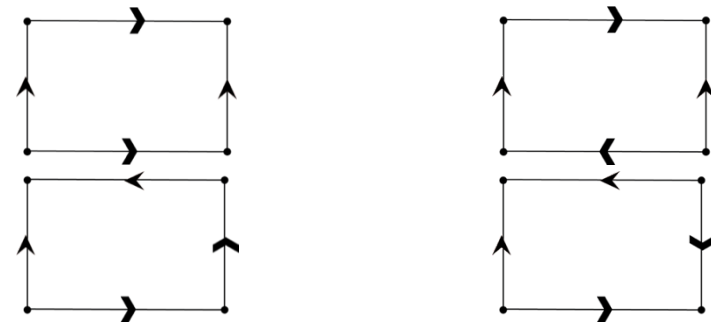

Figure 8. Adding two matching oriented edges

The next step would be to add two additional oriented edges to one of the configurations above.

In general, a polygonal arrow diagram is a polygon with directed edges with labels to mark the pairing that results from a finite number

of applications of the recursive definition above. Two figures represent the same arrow diagram when there is a bijection taking edges to edges that preserves labels, directions, and adjacency relations. Moreover, given a diagram, we can change the direction of both arrows of a couple that has to be identified.

Now to their semantics. A polygonal arrow diagram represents a closed (i.e., without boundary and compact) connected surface. Surfaces are two-dimensional manifolds. That is, they are topological spaces that are locally indistinguishable from the Euclidean plane.[14]

Arrow diagrams represent surfaces as quotient spaces of disks in the plane (all polygons are, in fact, topologically equivalent to circles). In order to specify precisely how to obtain such a quotient space, we can give intrinsic coordinates to each directed edge E_i as follows: 0 at the tail endpoint, 1 at the head endpoint; taking the length of E_i as unit, for each point p on E_i between the endpoints assign t to p, where t (between 0 and 1) is the distance of p from 0 (see Figure 9). Notice that the bottom-left vertex is now labeled 1 but will be labeled 0 if conceived as the starting point of the vertical left edge (all edges will be identified in groups). If E_1 and E_2 have matching labels, gluing the edges is, for all r such that $0 \leq r \leq 1$, identifying the points of E_1 with the points of E_2 with the same coordinate r.
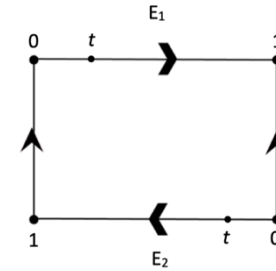

Figure 9. How to identify to edges

Given an arrow diagram, for each point p not on an edge, we identify p with itself and nothing else; for each point p on an edge, we identify p with itself and with q, where p and q are on paired edges and have the same coordinate. All the vertices are going to be identified in groups. The relation $x \sim y$ (x is to be identified with y) is an equivalence relation. We thus get a partition of the polygon P into a set of $\sim$ equivalence classes — that is, sets of the form $\{y \in P: x \sim y\}$.

These $\sim$ equivalence classes are the set-theoretic representatives of points of the surface Σ represented by the polygon P. A topological surface Σ is a set S of points together with a set of subsets of S, which are *the open sets* of the surface. Which subsets of S are the open sets of Σ? Let g be the function that sends each point p of the polygon P to the $\sim$ equivalence class to which it belongs — that is, $g(p) = \{y \in P: p \sim y\}$. A subset U of S is open in Σ if and only if $g^{-1}(U)$ is open in P (where a set of points of P is open in P if and only if it is the intersection of P with an open set in the Euclidean plane in which P is embedded). This is the standard definition of quotient spaces.

The two digons B_1 and B_2 in Figure 7 represent the sphere and the projective plane, which is a non-orientable surface (and therefore cannot be embedded in space). The diagrams in the first line of Figure 8 represent the torus (that is, a closed orientable surface of genus 1) and the *Klein bottle*, which is a non-orientable closed surface. The diagrams

---

14. A *closed connected surface* is a two-dimensional topological space in which (i) every point lies in an open set homeomorphic to the Euclidean plane (e.g., a disk minus its boundary), and (ii) for distinct points x and y there are disjoint open sets such that x lies in one and y lies in the other (that is, it has the property of being a Hausdorff space).

in the second line of Figure 8 can be reduced to B_2 and B_1, respectively, by gluing the left and top arrows together.

The last thing we have to show to establish the soundness of the proof passage of Theorem 1 is that the cutting and pasting operation leaves the represented surface invariant. To cut along a segment E running between two vertices, say v1 and v2, is (i) to give E a direction, say from v1 to v2, and (ii) to produce two new diagrams from the parts of the cut diagram on either side of E, which (a) preserves the edges of the pre-cut diagram with the new diagrams' labels and orientations and (b) gives to each new diagram a new edge (created by the cut) which is a copy of E, each directed to its copy of vertex v2 (see Figure 10).
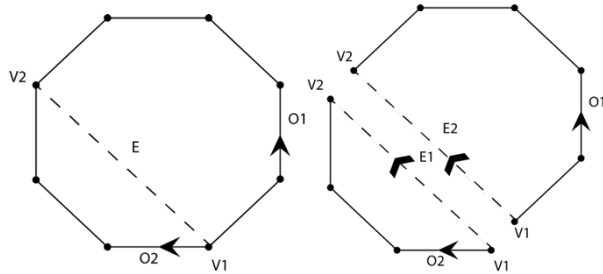

Figure 10. Cutting along an internal segment

We can then glue the two diagrams along another pair of matching edges, in this case O1 and O2 (see Figure 11). Doing so, we would obtain the same surface as the one obtained by gluing back directly E1 and E2. This is a consequence of the fact that the order in which we glue the matching edges does not matter.
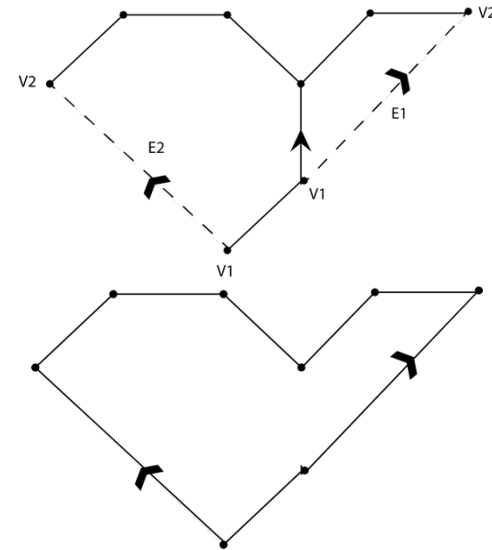

Figure 11. Gluing along an edge

Let me now turn to another operation. Surfaces admit the operation of *connected sum*. For example, performing the connected sum of two tori, we obtain a double torus. Intuitively, starting with the two surfaces, we subtract from each of them an open disk and then glue along the newly formed boundary circles (see Figure 12).
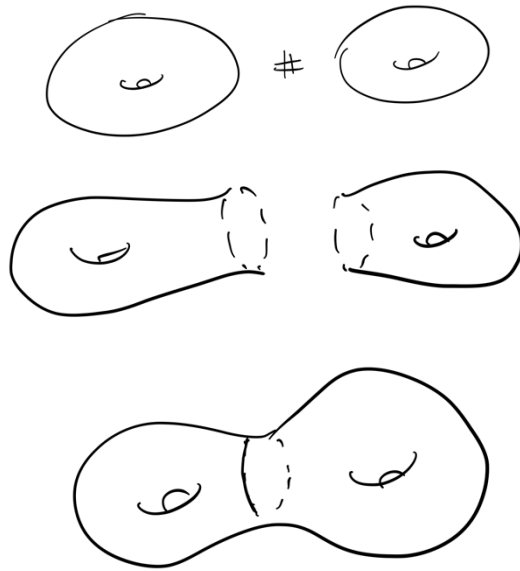
Figure 12. The connected sum of two tori

This operation corresponds to combining two arrow diagrams together. We insert one diagram on a vertex of the other. In Figure 13 we see how to connect two diagrams representing the torus — we imagine opening up the first diagram at the bottom right and the second diagram at the top left and then joining the two together to get the diagram shown in Figure 14.
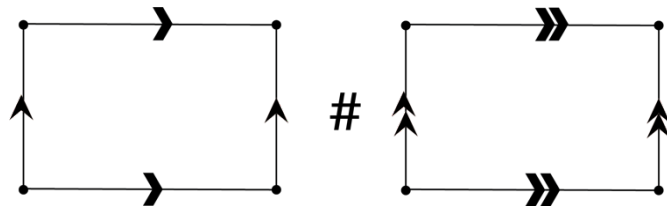

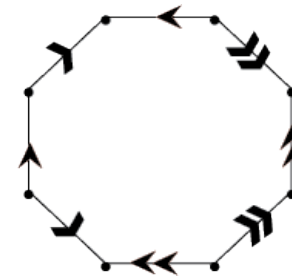Figure 13. Summing two diagrams

Figure 14. The diagram of a double torus

The system formed by arrow diagrams is both sound and complete with respect to its interpretation. Each well-formed diagram corresponds to a closed connected surface, and each closed connected surface corresponds to a well-formed diagram. That is, a topological space is such a surface if and only if it can be represented as a well-formed diagram. Performing the cutting and pasting manipulation and the manipulation corresponding to the connected sum, we always obtain other well-formed diagrams representing surfaces.

2.3 Diagrammatic Notational Systems

We saw that sequences of diagrams can be parts of proofs. This is because we can mentally manipulate them in ways that correspond to specific mathematical operations. These manipulations are cognitively simple and can be performed reliably by mathematicians. Crucially, arrow diagrams do not pose the risk of overgeneralization that is typical of geometric diagrams. Overgeneralization arises when a specific diagram is used to obtain general results. For instance, when we sketch a triangle to prove a result about all triangles, how can we be sure that we are not relying on features of our particular representations that are idiosyncratic to it? With arrow diagrams we do not have this problem because they form a sound system with respect to their

interpretation. No visible variable features of drawings of an arrow diagram, apart from the number of edges and their pairing and directions, are *topologically* relevant. Additional variations, such as variations of internal angles and of length ratios of diameters, are *geometrically* relevant — they do not matter for arrow diagrams, but they are relevant to geometric diagrams.[15] Hence, we are more liable to make a false geometrical generalization (than a false topological generalization) by relying unwittingly on a visible feature of a diagram that is not shared by everything in the range of generalization.

Second, arrow diagrams can be easily generalized. We saw that we can represent non-orientable surfaces — that is, surfaces that cannot be embedded in space. For example, the second diagram in Figure 8 represents the famous Klein bottle.[16] Furthermore, arrow diagrams representing surfaces generalize to solids. Henri Poincaré (1900) explains how to obtain three-dimensional spaces by identifying the faces of solid polyhedra. For example, the *Poincaré homology sphere* (i.e., a three-manifold whose homology groups are the same as the sphere, but it is topologically different from it) can be obtained by gluing the opposite sides of a solid dodecahedron as in Figure 15.
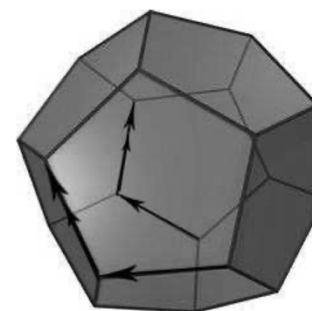

Figure 15. A diagram representing the Poincaré homology sphere

This is significant. Three-dimensional spaces cannot, in fact, be represented by something analogous to the sketches of surfaces; in order to see a three-dimensional space from the *outside* we would have to represent it in a four-dimensional space. The three-dimensional analogs of arrow diagrams for surfaces, however, are not so well behaved. When we consider the three-dimensional spaces resulting from gluing the faces of polyhedra, these are not necessarily manifolds because they may have singular points.

Let us return to arrow diagrams for surfaces. As we have seen, the fact that arrow diagrams can enter into the inferential structure of proofs can be explained by two facts: (i) they satisfy well-formedness conditions, and (ii) they are subject to precise rules of manipulation. In this respect, diagrammatic systems are on all fours with other mathematical notational systems.[17] It will be easier to think of physical instantiations of mathematical notations. We should keep in mind, however,

---

15. The distinction between geometric and topological features roughly tracks the one between *exact* and *coexact* features introduced by Kenneth Manders (2008) in the context of his analysis of Euclidean diagrams.

16. See Mark Colyvan (2012, 162) for a discussion of this point in relation to the efficacy of diagrammatic notations.

17. My discussion is an analysis of visual representations in the tradition inaugurated by Nelson Goodman's (1976) *Languages of Art* and developed by John Kulvicki (2003). However, my analysis is distinct from theirs because I work with mathematical notational systems, which have to satisfy more rigid constraints compared to other notational systems (and images more generally). In this respect, my work is more related to works focusing on logical notations; see Schlimm (2018).

that we can also imagine diagrams and other representations without writing them down. But what are mathematical notations? I will not answer this question in all its generality. Let me just remark that they are formed by elements, in our case, inscribed in a material medium that can be combined in different ways. Not all perceptual features of such elements carry mathematical content; as a result, they are not all relevant for their interpretation.[18] It is helpful, then, to distinguish between perceptual features that are *constitutive* from perceptual features that are merely *enabling*. For example, generally the color or width of the lines used in arrow diagrams or linear algebraic notations is merely enabling. In those notations, they cannot carry mathematical content. However, in the proof of the classification theorem for surfaces, the zigzagging has a specific mathematical meaning and is thus constitutive. In the statement of the theorem, colors are also constitutive since they track the pairs of sides to be glued together — although the same information is also conveyed by the labels. Crucially, the syntax of a notation cannot be altered by changing only enabling attributes. The partition of perceptual features of a notation into constitutive and enabling is given by the interpretation, which also determines which manipulations are mathematically meaningful.[19]

Arrow diagrams are an example of diagrams that are not merely a subjective representation but form a mathematical notational system. But further doubts concerning the legitimacy of diagrammatic systems in proofs could be raised. Problems might arise, for instance, if some of the constitutive features of a diagram were too difficult to identify or reproduce reliably. As I have argued elsewhere (De Toffoli 2022), notations have to satisfy three basic constraints:

1) a notation should be *cognitively accessible*: its constitutive perceptual features should be clearly identified, persistent, and stable;

2) a notation should be *reproducible*: it should be possible for an average practitioner to copy its constitutive perceptual features with relative ease and reliability, possibly with the aid of different tools such as a straightedge and/or computer; and

3) a notation should *support calculations* and/or inferences: it should be possible for an average practitioner to perform reasonably simple manipulations corresponding to mathematical operations.

In most cases of non-diagrammatic notations, such constraints are met as a matter of course. These constraints come in degrees. However, since we, as humans, share cognitive and perceptual limitations, there is a threshold that all notations have to meet in order to count as notations *at all*. The fact that such constraints have often been left implicit may have contributed to skepticism *vis-à-vis* the use of diagrams in proofs.

As Kenneth Manders (2008) has discussed at length with respect to Euclidean geometry, if exact metric properties had been constitutive features of Euclidean diagrams, then ancient Greek geometric practice would not have gotten very far. The resulting representations would not have been shareable and reproducible. That is why such exact metric features were only available from the text.[20]

The fact that arrow diagrams form a genuine notation is reflected in the fact that they present an algebraic interpretation.[21] We can label the

---

18. Even if I use the term *perceptual features,* in my analysis I want to include diagrams that are merely imagined. In this case, the perceptual features are the features that a physical instantiation of those mental diagrams would have.

19. For an in-depth discussion of the distinction between enabling and constitutive features of mathematical notations, see De Toffoli (2022).

20. Inspired by Manders's analysis of Euclidean diagrams, Brendan Larvor (2019) also spells out conditions that diagrams should satisfy to be used in proofs. Mine are similar in spirit but more general. See De Toffoli (2022) for a comparison between the two.

21. Similarly, Euclidean diagrams are used in systematic practices in a codified manner. Indeed, it is the possibility of codification that accounts for the

edges of an arrow diagram in Figure 5 and code it with the sequence, $a_1 b_1 a_1^{-1} b_1^{-1} a_2 b_2 a_2^{-1} b_2^{-1}$…. This can be useful, but transforming the diagrams into such *algebraic words* would obscure their topological meaning. The fact that arrow diagrams can be easily coded might give the impression that, after all, diagrams are mere abbreviations of other forms of display. But this is incorrect. First, the two proofs include diagrams and not codes. Eliminating the diagrams without substituting them with other representations would leave the arguments incomplete. That is, in those proofs, diagrams are *not* redundant. Second, although it is true that diagrams might be replaced by algebraic codes, it is far from obvious that such replacement would preserve the original proof. Before proceeding to argue for this last claim (Section 3), it will be fruitful to consider another example of how diagrams are used in mathematics — one that will highlight to an even greater degree the similarities between diagrammatic and non-diagrammatic mathematical notations.

2.4 Algebraic Diagrams

In the mid-twentieth century, after a long exile, diagrams began emerging once again in mathematical papers. This is because new types of diagrams were introduced in the field of homological algebra, namely commutative diagrams. I present an instance of a particular technique used extensively in this field: *diagram chasing*.[22] The mathematical technicalities involved in explaining the technique need not scare the reader off. What is important in what follows is not so much to have a firm grip on each mathematical detail but rather to form an idea of how these diagrams are used in proofs.

Roughly, we begin with a commutative diagram made of nodes and arrows, such as the one in Figure 16, in which the nodes are abelian groups, and the arrows are group homomorphisms.[23]
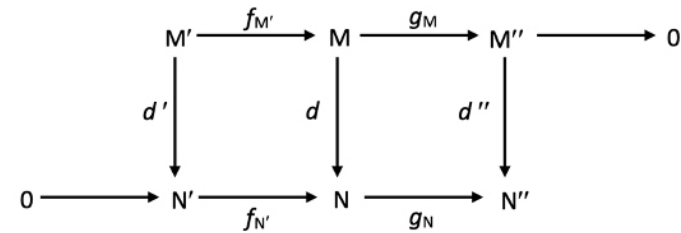

Figure 16. A commutative diagram

The fact that a diagram is commutative means that if we want to connect two nodes, it does not matter which arrows we follow: any two paths leading from the same starting point to the same endpoint are equivalent — in other words, the journey is precisely *not* the goal. For example, if we wish to connect M' to N in Figure 16, it makes no difference if we go right and down following arrows $f_{M'}$ and d or down and right following arrows d' and $f_{N'}$. Our goal will be to prove certain properties of the maps in a diagram (i.e., the arrows). If we focus on a given node, we can choose a specific element, say m' in M'. The technique of diagram chasing consists in moving such an element around the diagram by transforming it following the arrows, for example, moving m' along the arrow $f_{M'}$ and transforming it into the element m of the group M (where $m = f_{M'}(m')$). That is, we transform a given element belonging to one group in a node of the diagram by applying the homomorphisms corresponding to the arrows.

Each "0" in this diagram denotes a group with its identity element as sole member. This is a customary abuse of notation. In the text "0" denotes the identity of a group; additive notation is used.

---

existence of formal diagrammatic systems such as the one proposed by Jeremy Avigad, Edward Dean, and John Mumma (2009) for Euclidean geometry.

22. See Lang (2002) for the mathematical details and Feferman (2012) and De Toffoli (2017) for a philosophical discussion.

23. Note that here the arrows have a significantly different interpretation from that of the arrows in the diagrams for surfaces.

We say that a row of a commutative diagram is *exact* if, in that row, the image of one arrow is the kernel of the next. The *kernel* of an arrow (i.e., a homomorphism) is the inverse image of 0. That is, it is the set of all the elements that are sent to 0. Saying that the first row of Figure 16 is exact means two things: (1) Im $f_{M'}$=Ker g_M and (2) Im g_M = M'', this is because, by definition, the kernel of a homomorphism that sends all elements to 0 is the whole domain of that homomorphism. One last definition is needed. The cokernel of an arrow is the quotient of its codomain by its image. For example, looking again at Figure 16, we have Coker d' = N'/ Im d'.

I now present a passage of the famous *snake lemma* (named after the shape of the diagram it involves). In an introductory book on homological algebra, Charles A. Weibel writes, "We will not print the proof of the Snake lemma in these notes, because it is best done visually. In fact, a clear proof is given by Jill Clayburgh at the beginning of the movie 'It's My Turn'" (1994, 11).[24] Although I certainly agree that it is much easier to convey how the diagram chasing technique works with a live performance, the best I can do here is report a short passage of the proof.

> SNAKE LEMMA. Given the commutative diagram of abelian groups in Figure 17 in which the rows are exact, we can define the map:
>
> $$\delta: \text{Ker } d'' \to \text{Coker } d'$$
>
> with $\delta(z'') = h \circ f_{N'}^{-1} \circ d \circ g_M^{-1}$, where $h{:}N' \to \text{Coker } d'$ maps n' to $n' + \text{Im } d'$. And we get the following exact sequence:

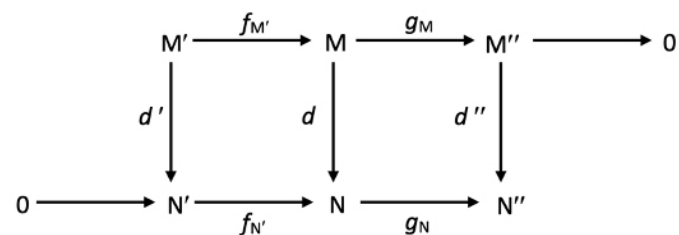$$\text{Ker}(d') \to \text{Ker}(d) \to \text{Ker}(d'') \to \text{Coker}(d') \to \text{Coker}(d) \to \text{Coker}(d'')$$

Figure 17. The snake lemma

The lemma is proved by diagram chasing. Each row is exact: the image of one arrow is the kernel of the next — e.g., Im $f_{M'}$ = Ker g_M. All rectangles commute. Figure 18 is an extended diagram in which the kernels and the cokernels of the vertical maps are displayed.
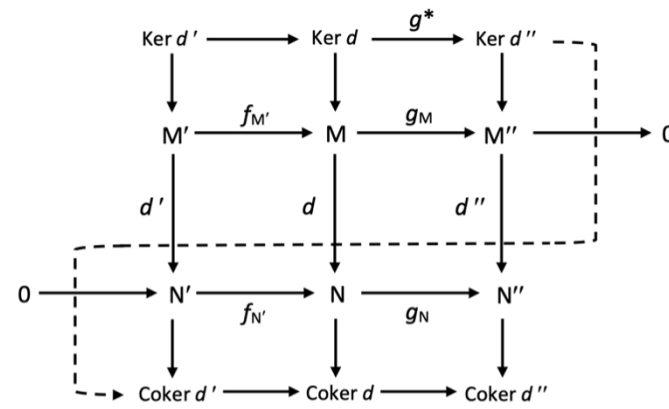

Figure 18. Extended diagram of the snake lemma

In this extended diagram, $g_*{:}\text{Ker } d \to \text{Ker } d''$ is just g_M restricted to Ker d (that is, it is the map induced by g_M on the kernels). The winding dashed arrow from Ker d'' to Coker d' represents the intended function δ. The downward

arrows from the kernels are inclusions (Ker d' is the set of $m \in M'$ that d' maps to 0_N; similarly for d and d''). Coker $d' = N' / \mathrm{Im}\, d'$ is the group of cosets of $\mathrm{Im}\, d'$ in N' ($\mathrm{Im}\, d'$ is $d'[M']$). The downward arrow to the cokernel of d' maps n' to the coset $n'+\mathrm{Im}\, d'$; similarly for d and d''. The middle two rows are exact, and all rectangles commute.

Following Serge Lang (2002, 158), I report only one step of the proof. However, I will prove this step in a much more detailed way, so as to make the argument intelligible to anyone with knowledge of basic group theory and to give some indication (in square brackets) of how it involves a *chase* around the diagram. Such information would normally be conveyed easily in a live lecture.

We define the homomorphism δ: Ker $d'' \to$ Coker d' in a roundabout way and show that

$$\mathrm{Ker}\,\delta \subseteq \mathrm{Im}\, g_*.$$

This has to be the case because of the exactness of the long exact sequence in the statement of the lemma (in fact, the strongest relation Ker $\delta = \mathrm{Im}\, g_*$ must hold). Facts from basic group theory are assumed without mention. Diagram chasing moves are indicated by comments in square brackets.

ONE STEP OF THE PROOF
Let z'' be any element of Ker d''. By exactness g_M: $M \to M''$ is surjective (as the final arrow $M'' \to 0$ has kernel the whole of M''). So, $\mathrm{Im}\, g_M = M''$. Now, $z'' \in$ Ker $d'' \subseteq M''$ (given). So, $z'' = g_M(z)$ for some $z \in M$.
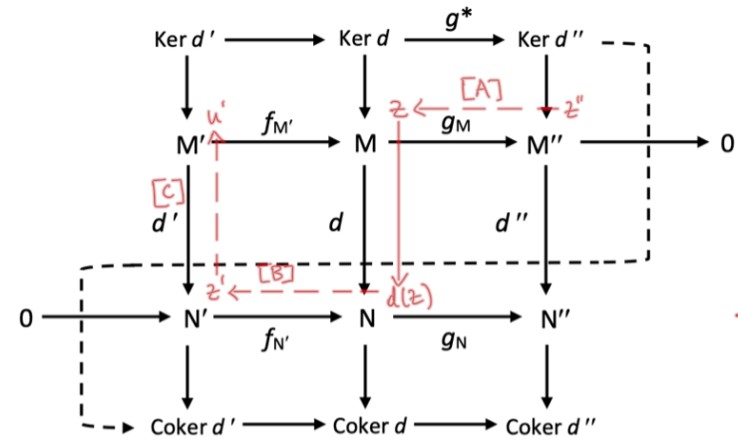

Figure 19. Diagram chasing

Now [A] we choose one such $z \in M$. [We move backward along g_M transforming z'' into z; see Figure 19.] As $z'' \in$ Ker d'', $0 = d''(z'') = d''(g_M(z)) = g_N(d(z))$ by commutativity (right middle rectangle), so $d(z) \in$ Ker $g_N = \mathrm{Im}\, f_{N'}$ by exactness. So, [B] for a unique $z' \in N', d(z) = f_{N'}(z')$. It is unique because $f_{N'}$ is injective (as the initial arrow $0 \to N'$ has image $\{0\}=$Ker $f_{N'}$ by exactness). [We move down along d turning z into $d(z)$ and then left.]

Now "let" $\delta(z'') = z' + \mathrm{Im}\, d'$. Why the scare quotes? Because z' is defined in terms of z, where z is just one choice from $g_M^{-1}(z'')$ [see A]. So, we need to show that a different choice from $g_M^{-1}(z'')$, say z_0, would make no difference – that is, that $z' + \mathrm{Im}\, d' = z_0' + \mathrm{Im}\, d'$. This holds and we can remove the scare quotes. For brevity we skip the argument.[25] We also skip the proof that δ is a homomorphism.

25. We have in effect defined δ as $h \circ f_{N'}^{-1} \circ d \circ g_M^{-1}$, where h: $N' \to$ Coker d' maps n' to $n' + \mathrm{Im}\, d'$.

Suppose now $z'' \in \text{Ker } \delta$, then $\delta(z'') = 0$ in Coker $d' = N'$ / Im d'. So $z' + \text{Im } d' = \delta(z'') = \text{Im } d'$. Hence $z' \in \text{Im } d'$; that is, [C] $z' = d'(u')$, for some u' in M'. Choose one such u' – it does not matter which.

By commutativity of the central left-hand square, we get $d(f_M(u')) = f_N(d'(u'))$. By [B] and [C], we also get $f_{N'}(d' (u')) = f_{N'}(z') = d(z)$. Therefore, $d(f_M(u')) = d(z)$. It follows that $0 = d(z) - d(f_M(u')) = d(z - f_M(u'))$. So, $z - f_M(u') \in \text{Ker } d$. [We can now move right with g_*.]

Recall that g_* is g_M restricted to Ker d. So $g_*(z - f_{M'}(u')) = g_M(z - f_{M'}(u')) = g_M(z) - g_M(f_{M'}(u'))$. Moreover, $f_{M'}(u') \in \text{Im } f_{M'} = \text{Ker } g_M$ by exactness; so $g_M(f_{M'}(u')) = 0$. Therefore, $g_*(z - f_{M'}(u')) = g_M(z)$ [previous two lines], $= z''$ [see A]. So $z'' \in \text{Im } g_*$. As z'' is an arbitrary member of Ker δ, Ker $\delta \subseteq \text{Im } g_*$.[26]

QED

As mentioned before, the presentation above is much longer than one we would find in a mathematics textbook or journal article. Many of the moves presented here as symbol manipulations are recognized immediately by practitioners from looking at the diagram, such as the identities established by commutativity of rectangles and exactness of rows.

Any presentation of this proof will make an appeal to exactness and to commutativity, and when such appeals are made, one needs to attend to the diagram to check that the identities claimed are in fact justified by those appeals. This point reveals that the visual thinking supported by the diagram involved in following the argument plays a much larger part than might appear, especially in relation to the expansive text. Moreover, visual thinking, where this includes both visually perceiving a diagram and visually imagining moves around the diagram, is essential to follow this argument.

What is not revealed by following a diagram chasing argument on its own is its cognitive efficiency relative to the mountain of pure symbol manipulation involved in following a diagram-free counterpart of the argument. Of course, one can follow a diagram-free counterpart stepwise, given sufficient indication of the prior information that is drawn on when inferring a new line, but one may still be in the dark about how individual steps contribute to larger sections of the argument and how these sections fit together, because the torchlight of attention tends to be focused narrowly on the individual steps. It is much easier to keep track of what is going on when following the diagram chasing argument, because attention to the diagram itself induces us to zoom out from individual steps (as well as to check identities based on exactness or commutativity). This relative cognitive efficiency applies not only to following an argument but also to discovering an argument. This becomes clearer as one becomes more practiced. It is true that different mathematicians have different preferences and modes of thinking; nevertheless, eliminating the diagrams from diagram chasing arguments would inevitably introduce some cognitive challenges.[27]

3. Diagrams That Are Not Abbreviations

If there were lingering doubts about whether diagrams can form notational systems, the example of commutative diagrams should have dispelled them. Proofs by diagram chasing show that diagrams do not necessarily threaten the validity of proofs.[28]

It is now time to tackle the second issue raised by Burgess: Are all diagrams, which are not mere illustrations, just abbreviations of more cumbersome linguistic expressions? The answer is no. It is true that diagrams can always be *coded* into linear displays. Not only can any digital image be encapsulated into a sequence of bytes and bits, but diagrams forming notations, which are discrete objects with a finite

26. Thanks are due to Marcus Giaquinto for helping me to make this proof more accessible and in clarifying its philosophical importance.

27. Thanks to one of the anonymous referees for this point.

28. See De Toffoli (2022) for a characterization and taxonomy of mathematical diagrams.

number of well-identifiable constitutive features, can be coded with system-specific encodings. For example, codes formed by algebraic words can be used for arrow diagrams for surfaces such as the one in Figure 5(a), which is coded with the sequence $a_1 b_1 a_1^{-1} b_1^{-1} a_2 b_2 a_2^{-1} b_2^{-1}$ …. Crucially, however, if we are supposed to leave the proof invariant when substituting diagrams with non-diagrammatic expressions, then it might be harder or outright impossible to eliminate diagrams. And if it is impossible, diagrams would be essential for *that* proof. This can happen in two distinct ways.

Diagrams are trivially indispensable to the proofs in formal diagrammatic systems. Sun-Joo Shin (1994) proved that *Venn*-II, a diagrammatic system she developed starting from Venn diagrams, is sound, complete, and equivalent to a system of monadic first-order logic. It is a tautology that without diagrams a formal *diagrammatic* system could not exist.[29]

The second way in which diagrams can be essential to proofs is at once subtler and more widespread. Diagrams can be essential without being indispensable inferential resources — that is, even if it is still possible to prove without them all results that are provable *with* them. They can still be essential in certain proofs in the sense that any diagram-free proof of the same result would be a *different proof*. This case requires a lengthier discussion, one that can offer guidelines for adjudicating between cases in which two proofs are the same and ones in which they are different.

3.1 Criteria of Identity for Proofs

Proofs are abstract objects. They can be considered equivalence classes of proof presentations. However, it is not always clear which equivalence relation is appropriate. At one extreme, we could consider all proofs of the same result to be equivalent. But this coarse-grained way of individuating proofs would be inappropriate to study mathematical practice. For instance, the common practice of proving the same result

in multiple ways would be unintelligible. Therefore, we should look for a more fine-grained criterion. As Marcus Giaquinto suggests, proof presentations containing the same "central idea" are generally considered to be equivalent, but

> if one's main concern is with what is involved in *thinking through* a proof, its central idea is not enough to individuate it: the overall structure, the sequence of steps *and perhaps other factors affecting the cognitive processes involved will be relevant*. (2008, 24, *emphasis added*)

There might be no fact of the matter about whether two presentations present the same proof or different proofs. Timothy Gowers (2007), a Fields medalist, admits in his popular blog that it was not only difficult to sharply characterize what it means for a proof to satisfy informal characterizations such as being explanatory, but "even the seemingly more basic question, 'When are two proofs the same?' was pretty hard to answer satisfactorily." It is plausible to think that there are no context-independent identity conditions for proofs.[30]

Criteria of identity for proofs are, in fact, determined only in a specific context of investigation.[31] More precisely, what has to be established to determine the criteria of individuation of a proof is the main *concern* or *purpose* for which individuation is required.[32]

For example, in one context, rigor may be the major concern; in *that* context, a presentation that makes totally explicit the assumptions, definitions, and logical steps on which the argument depends

---

29. See the discussion in Giaquinto (2008, 25) with respect to a formal diagrammatic system for Euclidean geometry.

30. This is also in line with Dawson (2006). Note that there might be canonical criteria of identity for *formal proofs*. These, however, would not easily generalize to traditional proofs.

31. Similarly, one could ask whether two performances are performances of the same symphony. Although the score is often recognized as what fixes the identity of musical pieces, how much variation is allowed will depend on the interest of the social group. This is a complex issue; see, e.g., Goehr (1992).

32. Note that the same individual or the same group of individuals can have different purposes in different times. What matters is the specific context of investigation.

presents a distinct proof from an informal exposition of the first.[33] But if the major concern is not about rigor but about *purity*,[34] hence about what mathematical concepts are involved, the degree of informality may count as a mere matter of presentation, and so a fully explicit presentation and an informal exposition could present the same proof.[35]

In the present context of investigation, the main concern is with the epistemic and cognitive benefits and drawbacks for someone working through a proof. In particular, understanding the argument, in the sense of having a grasp of what motivates the major lines of argument, so that the large-scale structure of the argument comes into view, is a major epistemic benefit. And cognitive efficiency, in the sense of the ease of working through the argument, is another significant benefit. For example, it works against efficiency if, in order to see why steps are valid, one has to hold in mind a heavy load of formulas on which later steps may depend or to search back frequently to find the immediate premises on which steps depend. Therefore, in *this* context, if one proof presentation is significantly better than another with regard to cognitive-epistemic pros and cons for someone working through them, the proofs presented are distinct proofs. This holds even if one of the presentations is obtained from the other by replacing all diagrams with non-diagrammatic encodings.

Let us consider the cases of arrow diagrams for surfaces discussed above. Imagine substituting all diagrams with algebraic codes in the proof of the classification theorem for surfaces presented in Massey (1991), a step of which is reported in Section 2.1. We can perform such

33. Note, however, that Jody Azzouni (2013) has argued that diagrammatic proofs can be "perfectly rigorous." See also De Toffoli (2021) for a discussion about the compatibility of rigor and diagrammatic reasoning.

34. See Mancosu and Arana (2015) for a discussion of purity in the context of mathematics.

35. Mathematicians can have two or even more concerns at the same time. For example, searching for a purely analytic proof of the Intermediate Value Theorem, Bernard Bolzano wanted to achieve both purity and rigor; see Detlefsen (2008b) and Kitcher (1975). I am thankful to one of the anonymous referees for this observation. My point is simply that privileging one concern over another might lead to the choice of different criteria of identity of proofs.

a substitution to represent the diagrammatic moves shown in Figure 6 — which I reinserted here as Figure 20 for ease of comparison.

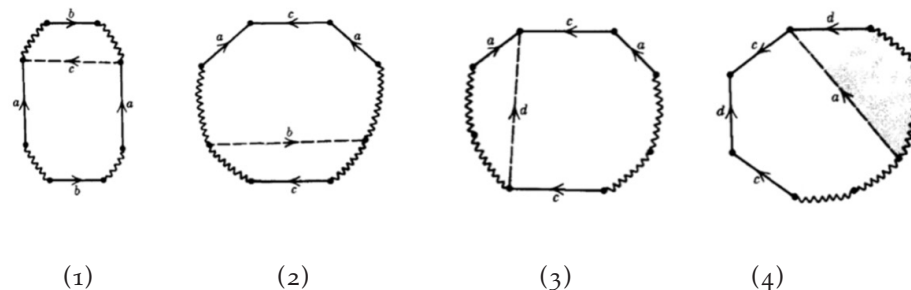(1)   (2)   (3)   (4)

Figure 20. A step of a proof involving arrow diagrams for surfaces

I will code the diagrams starting with their left edge and using Z as a convention for the zigzag notation that, as we saw above, is used to introduce generality into the diagrammatic representations. Start with the surface $aZbZa^{-1}Zb^{-1}Z$; now cut it so as to separate the surface ZbZ with c, a new edge (this is the first diagram), and glue the two new surfaces along b. We obtain $Zac^{-1}a^{-1}Zc$ (second diagram). Now, cut as to separate Za (third diagram) introducing the new edge d. We get $dc^{-1}d^{-1}Zc$, which is equivalent to $cdc^{-1}d^{-1}Z$ — which is part of the general pattern for orientable surfaces.

This brief presentation with algebraic words does not involve any diagram, but, as a matter of fact, in order to go through it and recognize it as valid, a mathematician would generally recreate (in the paper or mentally) arrow diagrams similar to the ones in Figure 20. There is, however, an alternative. We can spell out rules of manipulation of the algebraic codes themselves that correspond to the rules of manipulation of the diagrams. After listing such rules, we could check the correctness of the diagram-free proof without having to invoke any diagram whatsoever. But in this case, the diagram-free proof presentation does not present the same proof as the diagrammatic proof

presentation. This is because the two presentations differ with regard to epistemic and cognitive benefits and drawbacks. The diagrammatic proof presentation requires the reader to visualize specific spatial manipulations and recognize their validity. The diagram-free proof does no such thing. It is not only much easier to go through the diagrammatic argument, but it also allows us to grasp the proof more vividly. Moreover, the very subject matter of the two presentations differs: one is about surfaces, represented with arrow diagrams in an intuitive way, and the other is about algebraic words, which represent surfaces only indirectly. Arrow diagrams are needed to transition from algebraic words to surfaces since the words are codes for the diagrams and not for the surfaces themselves.

One could concede that this is true but downplay the importance of such a result by pointing at the fact that, after all, if we prove things about geometric or topological objects, it is not surprising that proofs involving geometric or topological representations would stand out compared to ones that do not involve them. There is a twofold answer to this complaint. First, the matter is not whether a result is or is not surprising, but rather whether it is true. Our question was: Can diagrams be essential to proofs? The example of the classification theorem for mathematical diagrams shows that the affirmative response is correct. Second, it is also possible to use the case of commutative diagrams, which are rarely about geometric or topological objects, to make the same point. Let us do that now.

The step of the proof of the snake lemma I presented invokes commutative diagrams. Once again, if we convert the proof into a diagram-free counterpart, a mathematician would generally reconstruct for herself the diagram in order to go through it.[36] Without the diagrams it is much, much harder to grasp the large-scale structure of the proof and what motivated the various steps. The diagrams represent in a compact and surveyable way the algebraic situation. They allow

36. Even when we implement or check a formally verified version of the snake lemma, we normally need to reconstruct the diagrams (again, either in paper or mentally).

mathematicians to identify viable proof strategies. Moreover, also for small-scale grasping, diagrams play a key role. They allow the reader to have a place to return to in order to check whether the algebraic manipulations are correct. Are those the right indices? Is it n or n'? What is the image of x under f? That the particular reasoning at play would change without the diagrams is also supported by the fact that, in jargon, it is common to refer to *diagram chasing* as a mathematical technique and as the hallmark of certain proofs. We need a diagram to perform the technique of diagram chasing!

4. Conclusion

Let me sum up the considerations of the previous paragraphs. In the case of proofs by diagram chasing, the diagrams are needed to grasp the large-scale structure of the proof and to help us carry out symbolic manipulations without having the burden of holding all indices in mind. In proofs by cutting and pasting arrow diagrams for surfaces, the diagrams enable us to use visualization in a rigorous way — it is this visualization that makes the diagrammatic proofs clearer and easier to verify. So, any transformation of the arrow diagrams proof of the classification theorem for surfaces and the commutative diagrams proof of the snake lemma into non-diagrammatic proofs will result in a significant loss of epistemic and cognitive benefits for anyone working through those proofs.

From the perspective of mathematical practice, a significant difference in the overall epistemic and cognitive advantages of presented proofs entails that the proofs themselves are different. This holds even if they have the same informational content. Therefore, from the perspective of mathematical practice, diagrams in the proofs of the classification theorem for surfaces and the snake lemma discussed earlier are essential to those proofs.

To return to the initial skepticism expressed by Burgess, while some illustrations are perhaps "no more essential to the proofs in whose midst they appear than the illustrations that accompanied many Victorian novels on first publication were essential to the literary value

of those works," this is in general not true for diagrams. Michael Detlefsen recognized this, in addition to recognizing the growing importance of diagrams in philosophy of mathematics:

> [T]he growing body of work on diagrammatic reasoning is of […] great interest and potential. It challenges the traditional ideas concerning the role of diagrammatic reasoning in proof and the development of mathematical knowledge more generally. It suggests, in particular, that diagrammatic reasoning has a justificative and not merely a heuristic role to play in proof. Much interesting work has already been done in this direction, and more is sure to follow. (2008a, 28)

In this paper, my goal has been to bring us a bit further in this direction.

Acknowledgments

References

Avigad, Jeremy, Edward Dean, and John Mumma. 2009. "A Formal System for Euclid's Elements." *The Review of Symbolic Logic* 2 (4): 700–768.

Azzouni, Jody. 2013. "That We *See* That Some Diagrammatic Proofs Are Perfectly Rigorous." *Philosophia Mathematica* 21 (3): 323–38.

Burgess, John P. 2015. *Rigor and Structure*. Oxford University Press.

Camp, Elisabeth. 2007. "Thinking with Maps." *Philosophical Perspectives* 21 (1): 145–82.

Carter, Jessica. 2018. "Graph-Algebras–Faithful representations and mediating objects in mathematics." *Endeavour* 42(2–3): 180–8.

——. 2019a. "Exploring the Fruitfulness of Diagrams in Mathematics." *Synthese* 196: 4011–32.

——. 2019b. "Philosophy of Mathematical Practice – Motivations, Themes and Prospects." *Philosophia Mathematica* 27 (1): 1–32.

Chemla, Karine. 2018. "The Proof Is in the Diagram: Liu Yi and the Graphical Writing of Algebraic Equations in Eleventh-Century China." In "Tools of Reason: The Practice of Scientific Diagramming from Antiquity to the Present," special issue, *Endeavour* 42 (2–3): 60–77.

Colyvan, Mark. 2012. *An Introduction to the Philosophy of Mathematics*. Cambridge University Press.

Davidson, Donald. 1978. "What Metaphors Mean." *Critical Inquiry* 5 (1): 31–47.

Dawson, John W., Jr. 2006. "Why Do Mathematicians Re-prove Theorems?" *Philosophia Mathematica* 14 (3): 269–86.

De Toffoli, Silvia. 2017. "'Chasing' the Diagram – The Use of Visualizations in Algebraic Reasoning." *The Review of Symbolic Logic* 10 (1): 158–86.

——. 2021. "Reconciling Rigor and Intuition." *Erkenntnis* 86: 1783-1802.

——. 2022. "What Are Mathematical Diagrams?" *Synthese* 200 (2): 1–29.

De Toffoli, Silvia, and Valeria Giardino. 2014. "Forms and Roles of Diagrams in Knot Theory." *Erkenntnis* 79 (3): 829–42.

Detlefsen, Michael. 2008a. "Proof: Its Nature and Significance." In *Proofs and Other Dilemmas: Mathematics and Philosophy*, edited by Bonnie Gold and Roger A. Simons, 3–32. Mathematical Association of America.

——. 2008b. "Purity as an Ideal of Proof." In *The Philosophy of Mathematical Practice*, edited by Paolo Mancosu, 179–97. Oxford University Press.

Feferman, Solomon. 2012. "And so on …: Reasoning with Infinite Diagrams." *Synthese* 186 (1): 371–86.

Ferreirós, José. 2016. *Mathematical Knowledge and the Interplay of Practices*. Princeton University Press.

Francis, George K., and Jeffrey R. Weeks. 1999. "Conway's ZIP Proof." *The American Mathematical Monthly* 106 (5): 393–99.

Giaquinto, Marcus. 2007. *Visual Thinking in Mathematics: An Epistemological Study*. Oxford University Press.

——. 2008. "Visualizing in Mathematics." In *The Philosophy of Mathematical Practice*, edited by Paolo Mancosu, 22–42. Oxford University Press.

Goehr, Lydia. 1992. *The Imaginary Museum of Musical Works: An Essay in the Philosophy of Music*. Clarendon Press.

Goodman, Nelson. 1976. *Languages of Art: An Approach to a Theory of Symbols*. 2nd ed. Hackett Publishing.

Gowers, Timothy. 2007. "When Are Two Proofs Essentially the Same?" *Gowers's Weblog*, October 4. https://gowers.wordpress.com/2007/10/04/when-are-two-proofs-essentially-the-same/.

Hahn, Hans. 1980. "The Crisis in Intuition." In *Empiricism, Logic and Mathematics: Philosophical Papers*, 73–102. Springer.

Kitcher, Philip. 1975. "Bolzano's Ideal of Algebraic Analysis." *Studies in History and Philosophy of Science Part A* 6 (3): 229–69.

Kulvicki, John. 2003. "Image Structure." *The Journal of Aesthetics and Art Criticism* 61 (4): 323–40.

Lang, Serge. 2002. *Algebra*. Rev. 3rd ed. Springer.

Larvor, Brendan. 2019. "From Euclidean Geometry to Knots and Nets." *Synthese* 196 (7): 2715–36.

Mancosu, Paolo, ed. 2008. *The Philosophy of Mathematical Practice*. Oxford University Press.

Mancosu, Paolo, and Andrew Arana. 2015. "Plane and Solid Geometry: A Note on Purity of Methods." In *From Logic to Practice*, edited by Gabriele Lolli, Marco Panza, and Giorgio Venturi, 23–31. Boston Studies in the Philosophy and History of Science. Springer.

Manders, Kenneth. 2008. "The Euclidean Diagram." In *The Philosophy of Mathematical Practice*, edited by Paolo Mancosu, 80–133. Oxford University Press.

Massey, William S. 1991. *A Basic Course in Algebraic Topology*. Springer.

Netz, Reviel. 1998. "Greek Mathematical Diagrams: Their Use and Their Meaning." *For the Learning of Mathematics* 18 (3): 33–39.

——. 1999. *The Shaping of Deduction in Greek Mathematics: A Study in Cognitive History*. Cambridge University Press.

Poincaré, Henri. 1900. "Deuxième Complément à l'analysis Situs." *Proceedings of the London Mathematical Society* 32: 45–110.

Radó, Tibor. 1925. "Über den Begriff der Riemannschen Fläche." *Acta Litt. Sci. Szeged*, 2: 101–121.

Schlimm, Dirk. 2018. "On Frege's Begriffsschrift Notation for Propositional Logic: Design Principles and Trade-Offs." *History and Philosophy of Logic* 39 (1): 53–79.

Shin, Sun-Joo. 1994. *The Logical Status of Diagrams*. Cambridge University Press.

Weibel, Charles A. 1994. *An Introduction to Homological Algebra*. Cambridge University Press.