Chapter 15
An Inquiry into the Practice of Proving in Low-Dimensional Topology

Silvia De Toffoli and Valeria Giardino

15.1 Introduction

Philosophy of mathematics has recently become more attentive to the practice of mathematics and in particular to the everyday work of mathematicians. One reason behind this "practical turn" is that mathematics should be acknowledged not only as an abstract science, but also as a human enterprise with its own dynamics that still need to be investigated in depth. Accordingly, it is common today to refer to the "philosophy of mathematical practice" (Mancosu 2008).

Setting our research in this context, our starting point is the analysis of some of the material representations used in the practice of mathematics. In Thurston's words, mathematicians

> use wide channels of communication that go far beyond formal mathematical language. They use gestures, *they draw pictures and diagrams*, and they make sound effects and use body language. (Thurston 1994, p. 166, *emphasis added*)

Among the wide variety of externalizations used by experts to convey and practice mathematics, only some of them are material and therefore easily shared, inspected and reproduced. Such material representations are introduced in a specific practice, and once they enter into the set of the available tools, they have in turn an influence on the very same practice. This process plays a significant role in

S. De Toffoli (✉)
Technische Universität Berlin, Institut für Mathematik, Straße des 17. Juni 136, 10623 Berlin, Germany
e-mail: toffoli@math.tu-berlin.de

V. Giardino
Département d'Etudes Cognitives, Ecole Normale Supérieure, Institut Jean Nicod (CNRS - EHESS - ENS), PSL Research University, Paris, France
e-mail: Valeria.Giardino@ens.fr

© Springer International Publishing Switzerland 2015

G. Lolli et al. (eds.), *From Logic to Practice*, Boston Studies in the Philosophy and History of Science 308, DOI 10.1007/978-3-319-10434-8_15

the practice of mathematics. One of the aims of this article is to investigate the conditions for its manifestation. Moreover, we will analyze what kind of cognitive abilities are triggered by the use of pictures in low-dimensional topology. In fact, we claim that a specific kind of imagination comes into play when dealing with visual representations in this field, which we label *manipulative* imagination. This notion will be used to characterize what it means to "see" in topology. We will propose that reasoning in low-dimensional topology is based on preexisting cognitive capacities—mathematicians imagine a series of possible manipulations on the representations they use—and is modulated by expertise: representations are cognitive tools whose functioning depends in part from preexisting cognitive abilities and in part from specific training.

Moreover, the actual practice of proving in low-dimensional topology involves a kind of reasoning that cannot be reduced to formal statements without loss of intuition. In this sense, visualization plays a specific epistemic role in this practice. We will show several examples of reasoning which are representationally heterogeneous, i.e. neither entirely propositional nor entirely visual.[1] This form of reasoning is shared by experts: it is the kind of reasoning that one has to master to become a practitioner. Moreover, the manipulations allowed on the representations—what we will define as *permissible* actions—as well as the representations themselves are epistemologically relevant. This is because they are integral parts both of the reasoning and of the justification provided. Inferences involving visual representations are permissible only within a specific practice and in this sense context dependent: this leads to the establishment of local criteria of validity.

In Sect. 14.2.2, we will introduce simple examples of reasoning in topology involving pictures and text in order to make the reader acquainted with various representations. In Sect. 14.4.1, we will focus on a specific proof in low-dimensional topology: Rolfsen's demonstration of the equivalence of two presentations of the Poincaré homology sphere. To do so, we will introduce the mathematical background only to the extent needed in this context. In Sect. 14.4.2, we will analyze our case study. We will discuss the form of topological arguments, in particular the role of pictures in topological reasoning. Moreover, we will consider how the notion of "seeing" in topology depends on our spatial-motoric intuitions of three-dimensional space and can be characterized using the notion of *manipulative* imagination. Lastly, we will analyze the reliability of this specific practice. In Sect. 14.4.3, we will sum up our conclusions and present possible lines of further research.

[1]Examples of representationally heterogeneous arguments can be found also in other areas. For example, Shin (2004) describes the project of funding a heterogeneous logic by Barwise and Etchemendy (1996) as a very fruitful one.

15.2 Reasoning in Topology

Topology is a branch of geometry that focuses on qualitative properties of objects, while ignoring quantitative ones. In order to represent a topological object, we have to choose one of its particular geometric shapes. Low-dimensional topology focuses on the study of objects of dimension four or less. It is particularly interesting to investigate the practice of proving in this subfield because, as we will see, it is deeply influenced by our intuition of space.

Let us start with an example in dimension two. Representations of surfaces can be "manipulated in space" by exploiting our familiarity in manipulating concrete objects, as if the objects of topology were made of modeling clay. For example, in topology the surface of a cup and that of a doughnut are equivalent: they are both homeomorphic to a *torus*. (In topology, objects are considered up to homeomorphism, i.e. continuous transformations whose inverse is also continuous.) To prove this, we have only to exhibit an appropriate deformation that takes the surface of a cup into the one of a doughnut. Students are trained to "see" transformations such as this and to move freely among different geometric shapes of the same topological object without need to justify these equivalences in other ways. More sophisticated arguments could be used by introducing the machinery of algebraic topology, but this is often not requested by the practice.

For example, the *torus* can be defined as a square with its sides identified. In order to explain how this is possible, let us first analyze a simpler example. Given a square with boundary, that is, a surface homeomorphic to a disk $\mathbb{D}^2$, we glue (i.e. identify) two of its opposite sides in order to obtain another surface. If the two sides are glued in the same direction, as indicated by the arrows in Fig. 15.1a, we will obtain a *cylinder* (Fig. 15.1b). If the two sides are glued in the opposite direction, as indicated in Fig. 15.2a, we will obtain a *Möbius band* (Fig. 15.2b).

In order to obtain the torus from a square, we identify all its four sides in pairs. First, we identify two of them in the same direction, as in the case of the cylinder

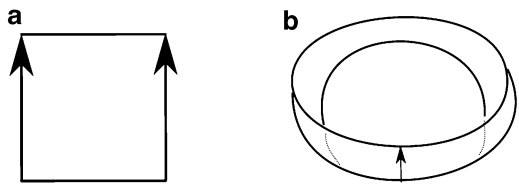

Fig. 15.1 Constructing the cylinder

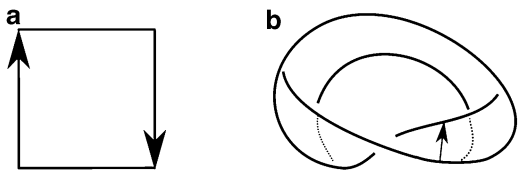

Fig. 15.2 Constructing the Möbius band

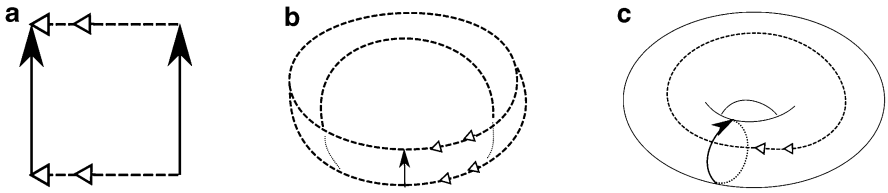

Fig. 15.3 Constructing the torus

(Fig. 15.3b), and then the other two again in the same direction (Fig. 15.3c).[2] In Fig. 15.3c, one can see the torus with two marked curves, where the gluings, i.e. the identifications, were made.

In discussing the role of notation in mathematics, Colyvan takes into consideration these square diagrams with arrows indicating the gluings:

> This algebraic topology notation is something of a halfway house between pure algebra and pure geometry. It is both notation and a kind of blueprint for construction of the objects in question. The first seems to belong to algebra, while the second is geometric. But whichever way you look at it, we have a powerful piece of notation here that does some genuine mathematical work for us. (Colyvan 2012, Ch. 8)

This visual presentation of the torus can be formalized as a quotient of the unit square $\mathbb{D}^2/\sim$. We consider the square in a coordinate system such that the edges are situated in the points $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$ and the equivalence relation $\sim$ identifies points (x, y) and (x', y') according to the following (see Fig. 15.3a):

$$(x, y) \sim (x', y') \Leftrightarrow \begin{cases} (x, y) = (x', y') \text{ or} \\ \{x, x'\} = \{0, 1\} \text{ and } y = y' \text{ or} \\ \{y, y'\} = \{0, 1\} \text{ and } x = x' \end{cases}$$

Without visualizing the transformations from Fig. 15.3a to Fig. 15.3c, which are simplified by the particular notation, one can hardly topologically understand what the above formula defining the equivalence relation is about. In this sense, visualization is essential to the specific kind of understanding proper to topology. The same holds for more complex examples.

For instance, consider, the decomposition of the 3-sphere (the equivalent of the sphere in one more dimension) as the union of two solid tori.[3] Jones uses this

[2] If we identify two sides in the same direction and the other two in the opposite direction, we obtain the *Klein bottle*.

[3] A solid torus is a torus that is filled. While a torus is a surface without boundary homeomorphic to $\mathbb{S}^1 \times \mathbb{S}^1$ (i.e. the product of two circles), a solid torus is homeomorphic to $\mathbb{D}^2 \times \mathbb{S}^1$ (i.e. the product of a disk and a circle), and its boundary is a torus.

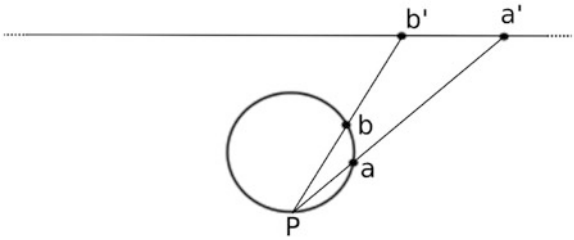

Fig. 15.4 A circle is homeomorphic to a line plus a point

example to emphasize the importance of visualization in low-dimensional topology. He compares a visual presentation with the following formal one: $\mathbb{S}^3 = T_1 \cup T_2$, with

$$T_1 = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 : x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1, x_1^2 + x_2^2 \leq 1/2\}$$
$$T_2 = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 : x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1, x_3^2 + x_4^2 \leq 1/2\}$$

Let us now "see" how the 3-sphere can be filled up by two solid tori beginning with two preliminary remarks.[4]

1) The 1-sphere $\mathbb{S}^1$ can be decomposed into a line $\mathbb{R}$ and point at infinity: Fig. 15.4 illustrates that each point in a circle except one (point P) can be put in correspondence to a point in a line. Point a goes to the point a', b to b', etc. Then, point P will be sent to infinity. Through this map, we get an homeomorphism between the circle $\mathbb{S}^1$ and the line $\mathbb{R}$ with a point at infinity added to it $\mathbb{R} \cup \infty$. Analogously, a sphere $\mathbb{S}^2$ can be obtained from a plane $\mathbb{R}^2$ by adding a point at infinity; and the 3-sphere $\mathbb{S}^3$ can be obtained from $\mathbb{R}^3$ by adding a point at infinity.

2) Given a line, we can rotate it around one of its points, e.g. the origin, and obtain the plane. We can also take a half-line, the one with the origin as its endpoint, and still obtain a plane after the same rotation. Similarly, we can start with a plane or a half-plane and rotate it around a line. For example, the xy plane in the standard coordinates if rotated around the y axis gives rise to the three-dimensional space $\mathbb{R}^3$.

In Fig. 15.5, the 3-sphere is represented as $\mathbb{R}^3 \cup \infty$ and $\mathbb{R}^3$ is the result of the rotation of the plane of the paper (the xy plane) along $\mathbb{R}_y$, the y axis. Let $\mathbb{D}_1$ be a disk (with boundary) in the half-plane $x < 0$; after the rotation of the plane around $\mathbb{R}_y$, it will form a solid torus, intersecting the xy plane in another disk: $\mathbb{D}_2$. In this way we get a representation of the 3-sphere with a solid torus inside it. Now, in order to prove that the 3-sphere can be decomposed as the union of two solid tori, we want to prove that also the complementary space of this solid torus is a solid torus.

Each of the segments depicted in Fig. 15.5 connecting a point of the boundary of $\mathbb{D}_1$ to the corresponding point of the boundary in $\mathbb{D}_2$ will give rise to a disk after the rotation. These disks can be parametrized by the line $\mathbb{R}_y$ plus the point at infinity since each intersects $\mathbb{R}_y \cup \infty$ in its midpoint (we can choose all these segments to intersect the circle $\mathbb{R}_y \cup \infty$ orthogonally). Thus, the complementary space is homeomorphic to $\mathbb{D}^2 \times \mathbb{S}^1$, a disk times a circle, which is a solid torus.

[4]The following argument can be found in Fomenko (1997, pp. 123–124).

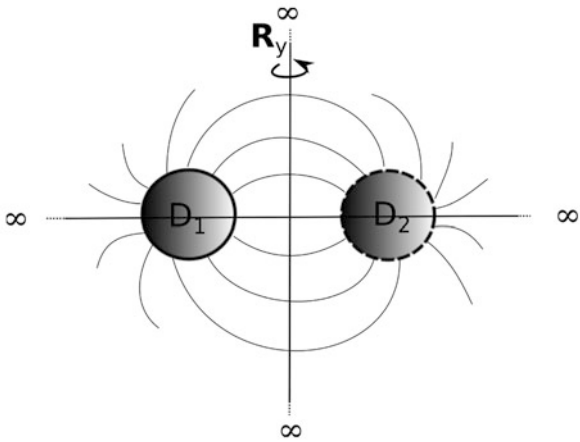

Fig. 15.5 Decomposition of $\mathbb{S}^3$ in two solid tori

Jones, considering a formal and a visual way of presenting the 3-sphere as the union of two solid tori, claims that

> [the] formal picture [. . .] is complete but inadequate. If one does not 'see' the other picture [. . .] one is not ready to take the next step in low-dimensional topology. Of course this is just the beginning. There are more complex things to 'see' and *sequences* of such visions are compounded one upon another *in the same way as the elementary logical steps in a formal argument*. If one 'sees' the pictures, then one understands, but otherwise one cannot follow. In principle one could formalize the whole argument, but *that would add nothing*. (Jones 1998, p. 213, *emphasis added*).

In Jones' view, an argument in this domain can be broken down into units, no matter if they are elementary logical steps or pictures. As we will later see, a formalization of an argument whose units are pictures would often not be relevant for the practice. This is characteristic of topology in general: often such formalizations would hide the relevant (topological) reasoning that is externalized by the pictures.

In the following, we will illustrate a proof in topology as an example of an argument composed of a sequence of pictures plus textual instructions on how to interpret them.

15.3 Rolfsen's Proof

We will now present the core of our case study: a proof of the equivalence of two presentations of the Poincaré homology sphere, taken from a popular graduate textbook: *Knots and Links* by Rolfsen (1976). The first presentation of

this 3-manifold[5] is a *surgery code*, while the second one is a *Heegaard diagram*. In order to explain this proof, we will briefly define the representations and techniques used to obtain them.

15.3.1 Dehn Surgery and Heegaard Diagrams

First, let us consider *Dehn surgery* on the 3-sphere. To do so, we introduce mathematical *knots*, which are smooth simple closed curves in the 3-sphere.[6] Knots are considered up to smooth deformations, i.e. we are not interested in their specific geometric shape, but in the way they are knotted.

In *Dehn surgery*, first we take a knot in the 3-sphere, then we thicken the knot to a tube in order to obtain a knotted solid torus. Next, we cut it out from the 3-sphere and glue it back in a different way to obtain another 3-manifold. The way in which the tube is glued back in can be coded by a rational number. Any closed compact 3-manifold can be obtained by Dehn surgery on some knot, i.e. it can be coded by a knot plus a rational number.[7] For example, a code for the Poincaré homology sphere is represented in Fig. 15.6.

To give a hint of the reasons why Dehn surgery is possible at all, consider an analogous process in lower dimension: in Fig. 15.7 we start with a circle (i.e. a 1-sphere, which corresponds to the 3-sphere in our example) and two points on it (i.e. a 0-sphere, which corresponds to the knot, i.e. a 1-sphere); then, we thicken these two points (this corresponds to thicken the knot) and get two segments (instead of a tube). After that we cut them out from the circle and glue them in a different way. We obtain two circles, which is a different topological manifold from one circle, the manifold we started with.

Let us now consider another way to present manifolds: *Heegaard diagrams*. To do so, we introduce *handlebodies of genus g*, which are balls with g handles attached. For example, a solid torus is, a handlebody of genus 1. A handlebody of genus g has g holes.

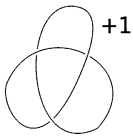

Fig. 15.6 The surgery code for the Poincaré homology sphere

[5] A 3-manifold is the equivalent of a surface in one more dimension. It is a three-dimensional topological space which is locally homeomorphic to $\mathbb{R}^3$, the Euclidean space.

[6] In a previous study, we considered the use of knot diagrams in relation to knot types. We claimed that the key feature of knot diagrams is their "dynamicity": experts manipulate them according to different sets of possible transformations (De Toffoli and Giardino 2014).

[7] See Fomenko (1997, Ch. 9) for details.

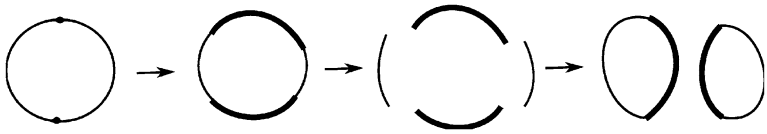

Fig. 15.7 Surgery in lower dimension

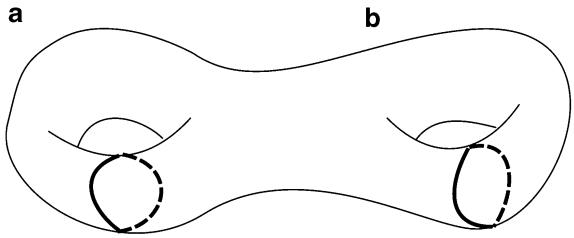

Fig. 15.8 A system of meridians of a genus 2 handlebody

It is possible to construct closed compact 3-manifold by gluing two handlebodies H_1 and H_2 of the same genus along their boundaries ∂H_1 and ∂H_2, which are two orientable closed surfaces of the same genus.

Let us focus on the case of genus two, which is the one that interests us for the proof which we will later analyze. To describe the gluing, i.e. a homeomorphism between the two boundaries, we consider a and b, the meridian loops of the first handlebody, which are two curves in ∂H_1, and the boundary of the first handlebody, as in Fig. 15.8. Then, in order to specify the result of the gluing, it is enough to know $f(a)$ and $f(b)$, the images of these curves under the attaching homeomorphism on ∂H_2, the boundary of the second handlebody. So, all we need to know in order to construct a 3-manifold from two handlebodies of genus two is a pair of simple closed curves in the boundary of a handlebody (these curves are interpreted as coding the gluing, since they are the images of the meridian loops of the first handlebody on the surface of the second.) This information, however, cannot be so easily conveyed, since it would require drawing curves on pictures of three-dimensional objects. To overcome this problem, Heegaard diagrams are introduced, which are two-dimensional diagrams containing all the relevant information. Their two dimensionality makes the presentation easier to draw and can be effective in a sense that we will explore later.

To construct a Heegaard diagram, we imagine cutting open ∂H_2, the boundary of the second handlebody, so that it lays flat on the plane and then we trace the image under this transformation of the curves determining the gluing (these curves were on ∂H_2 and thus will still be represented on this "flat" presentation of ∂H_2). First, we consider the meridian loops of ∂H_2 (call them A and B); then, we cut the surface ∂H_2 along them in order to create a surface with four circles as boundary (for genus g, we will have $2g$ circles), as in Fig. 15.9a. This is equivalent to a sphere with four holes, as in Fig. 15.9b: we imagine "inflating" the object represented in Fig. 15.9a so that it becomes a sphere.

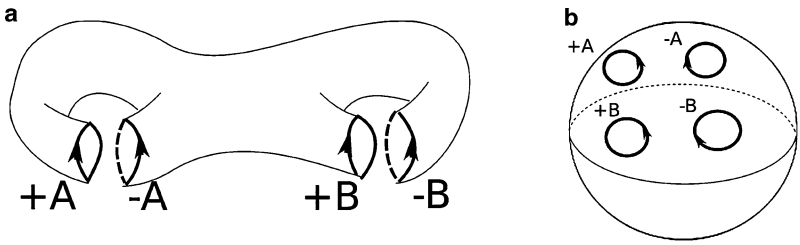

Fig. 15.9 Constructing a Heegaard diagram

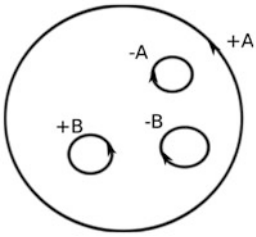

Fig. 15.10 "Flat" presentation of the boundary of a genus two handlebody

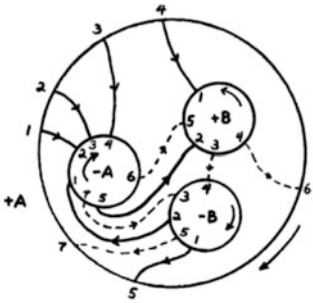

Fig. 15.11 The Heegard diagram taken from Rolfsen representing the Poincaré homology sphere

Then, from one boundary hole, we stretch the surface onto the plane, obtaining a disk with three holes. In Fig. 15.10, we stretched from the hole labeled $+A$.

While applying these transformations, we have to keep track of where the curves $f(a)$ and $f(b)$ defining the gluing are going, i.e. their image under this transformation. After the modification of the boundary of the second handlebody, these curves will be lying in the planar disk with holes. Then, a Heegaard diagram of genus two is a diagram of a disk with three holes endowed with a set of curves; see Fig. 15.11 for an example. Note that the set of lines in this diagram is actually to be interpreted as two closed curves, since the boundary circle are pairwise identified: $+A$ with $-A$ and $+B$ with $-B$, the signs denoting the opposite orientations which we need in order to glue them correctly.[8] In particular, there is a solid curve and a dotted one. Let us follow the solid one. We can start from $(1, +A)$, the point marked 1 in the external circle $+A$; then, we reach 2 in $-A$, but $-A$ is identified

[8]See Fomenko (1997, Ch. 5) for details.

with $+A$, so we arrive at 2 in $+A$; then for similar reasons, we reach in the order $(3,-A) = (3,+A)$, $(4,-A) = (4,+A)$, $(1,+B) = (1,-B)$, $(5,+A) = (5,-A)$, $(2,+B) = (2,-B)$, $(1,-A) = (1,+A)$. At the end we have returned at the starting point: the curve is indeed closed.

The information about these closed curves, so encoded, is enough to uniquely determine the gluing of the two handlebodies.

15.3.2 Two Presentations of the Poincaré Homology Sphere

In 1900 Poincaré, in the second supplement to his *Analysis Situs* (Poincaré 1900), had conjectured that a compact 3-manifold with the same homology groups as the 3-sphere would be homeomorphic to it. Five years later, he gave a counterexample constructing a manifold now known as the *Poincaré homology sphere*. This is a compact 3-manifold with the same homology as the 3-sphere but not homoeomorphic to it, because its fundamental group[9] is not trivial.

In Fig. 15.12, we present Rolfsen's direct proof of the equivalence of a surgery code and a Heegaard diagram of the Poincaré homology sphere by reporting the pictures and the accompanying text with the instructions on how to interpret them.[10] This proof (Rolfsen 1976, pp. 249–250) is accepted as a valid argument and, to our knowledge, is the only direct proof of the equivalence of these two presentations. We will discuss this proof in the next section.

We start with Dehn surgery, considering a trefoil knot with associated rational number $+1$; see Fig. 15.6. In the first picture of the proof, we see a thick trefoil forming a knotted tube with a curve indicating Dehn surgery to which another tube with a curve has been added.

The main idea of the proof is to continuously deform this object in order to obtain another known presentation of the same 3-manifold. The transitions between the first four pictures are topologically interpreted as homeomorphisms. While applying these transformations, we are going to keep track of the two curves. Training is normally required to follow these deformations, but we nevertheless hope to convey an idea of what a proof in topology might look like (it is not necessary to be able to follow all the transformations in order to get the gist of the proof).

We have to follow the instructions given in the text in order to identify the various transitions connecting the pictures. For example, to move from the fourth to the last picture, as the knife suggests, we cut the surface open and lay it flat on the plane. This is "almost" a Heegaard diagram: we just have to pick one of the holes and stretch it "outside" in order to obtain one. If we choose the hole labeled $+A$, we get the Heegaard diagram in Fig. 15.11.

[9]The fundamental group is a very important algebraic invariant used to study the shape of topological spaces. See Massey (1991, Ch. 2) for details.

[10]We sincerely thank Professor Rolfsen who gave us permission to reuse the material from his book.

The argument is contained in the series of pictures[...] Starting with Dehn's example, we drill out a tube from the trefoil's exterior and add it to the $\mathbb{S}^1 \times \mathbb{D}^2$, making the latter into a handlebody of genus two which is to be attached to the outer part of (1) by the characteristic curves shown. The isotopic deformation (1) - (4) shows that this outer part is also a handlebody. (Since the deformation is done equivariantly with respect to 180° rotation about the vertical, the back of the figure looks exactly like the front). Again the solid and the dashed curves of (4) are directions for sewing a handlebody onto the handlebody on the outside of the surface pictured. Cutting this surface open and flattening with the the inside toward us gives us the pictures (5) in the plane (plus a point at ∞). Moving the point at infinity to be in the hole labelled +A gives us exactly the Heegaard diagram of Poincaré's example.

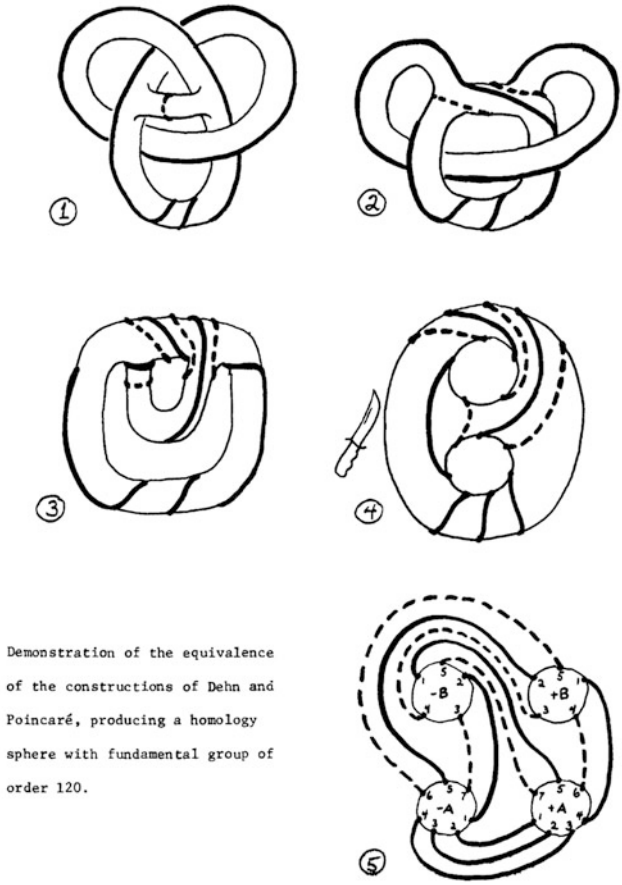

Fig. 15.12 Rolfsen's proof (Rolfsen 1976, pp. 249–250)

15.4 Discussion of the Case Study

What is striking about Rolfsen's proof is that even if it is clearly very far from a formal proof, and also far from a non-formal proof written in natural language, it is nonetheless (accepted as) valid. In this specific example, a formal description of the topological deformations would be beside the point.[11]

However, this should not be surprising, since, as Thurston writes,

> ...we should recognize that the humanly understandable and humanly checkable proofs that we actually do are what is most important to us, and that they are quite different from formal proofs. For the present, formal proofs are out of reach and mostly irrelevant: we have good human processes for checking mathematical validity. (Thurston 1994, p. 171)

The question is: which "human processes" would make mathematicians conclude that this proof is valid? This issue is also crucial to determine what makes low-dimensional topology so special as a subfield of mathematics.

In the following, we consider some of the features of this proof.

15.4.1 Heterogeneous Arguments

One evident feature of Rolfsen's proof is that it consists of pictures *and* text. The sequence of pictures is accompanied by a set of instructions given in natural language. Without the pictures, it would be impossible to understand the text; conversely, without the text, it would be very hard to correctly interpret the pictures: neither one is complete without the other. The argument requires both for its cogency.[12] Thus, the proof is not purely visual, for the same reason for which it cannot be considered as purely syntactic: it is representationally heterogeneous. In order to understand it, one has to be guided by the text on how to imagine the continuous transformations whose discrete steps are represented by the pictures.

The interdependence of pictures and text is not an exclusive feature of proofs in low-dimensional topology. Manders (2008) has discussed in depth how diagrammatic and propositional content are interacting and are both essential in the case of Euclidean geometry. Nevertheless, there are some differences between this proof (and others) in topology and most proofs in Euclidean geometry. For example, in the case of Rolfsen's proof, once the context, which is linguistically defined, is understood and the goals are set, then the text becomes superfluous

[11]This is not an isolated case. Similar phenomena can be observed also in research articles, for example, in the subfield connecting low-dimensional topology and hyperbolic geometry. See, for reference, Adams and Reid (1993).

[12]In our view, pictures do not always have to be physically drawn: it is sometimes sufficient to imagine them. Nonetheless, we will also argue that the fact that they can be physically drawn has cognitive advantages, since it makes inspection and reconfiguration easier and is the condition for sharing content with other practitioners.

for our understanding. That is to say that once we know what *to do* with the pictures—and this can only come along with expertise—we can concentrate on them exclusively. As Sullivan (2013) writes about this same proof: "you can't say you really understand the proof until you reach the stage where the pictures alone would suffice." The only place where we would still need instructions in words would be to go from the fourth to the fifth picture, but in this case, the instruction "cut" is substituted by the icon of a knife. This does not happen in Euclidean proofs, where the text guides us through the interpretation of the diagrams step by step and gives us necessary information that the picture alone does not provide (the text describes the articulated construction required to draw and work with Euclidean diagrams).

Let us now return to the argumentative form of Rolfsen's proof. Recall Jones' quotation in Sect. 15.2, where he claimed that "sequences of [...] visions are compounded one upon another in the same way as the elementary logical steps in a formal argument." Accordingly, we have seen that the majority of the steps into which Rolfsen's proof can be broken down are transformations that lead from one picture to another, as the text giving the instructions on how to interpret the pictures clarifies. Therefore, the representations give a material form to the transformations (and in this sense they "externalize" them) because they allow experts to perform "epistemic actions" (and in particular to draw inferences) on them. By *epistemic actions*, we mean actions that are performed with an epistemic aim.[13] Moreover, these actions are controlled by the shared practice: the set of legitimate transformations is limited and determined by the context.

As Larvor explained in a recent article by comparing different forms of mathematical arguments:

> ... if an argument includes an inferential action that manifests or manipulates the subject-matter, or a representation thereof, then formalising this argument in a general logical language must either misrepresent or fail to include this action. Moreover, we can say something in the direction of explaining how informal arguments work as arguments: they are rigorous if they conform to the controls on *permissible actions* in that domain. (Larvor 2012, p. 724, *emphasis added*)

Through an inferential action, it is possible to manipulate the representation itself for epistemic purposes. In Rolfsen's proof, the inferential actions consist in manipulating pictures; for this reason, any formalization would fail to capture the inferential actions performed. Moreover, we can proceed from one representation to another by applying actions that are permissible, i.e. allowed and controlled by the shared practice of the subfield. These epistemic actions are inferential actions: they constitute the units of a mathematical argument.

The commented pictures in Rolfsen's proof—*in sequence*—are thus the argumentative form of the proof. The representations constituting the argument are heterogeneous and yet adequate to its mathematical context: pictures thus play a

[13] *Epistemic actions* have been characterized by Kirsh and Maglio (1994) as "actions performed to uncover information that is hidden or hard to compute mentally" in contrast to *pragmatic* actions, which are "actions performed to bring one physically closer to a goal."

relevant epistemic role. This is in line with the actual practice of mathematics, as the first words of the proof makes clear: "The argument *is contained in* the series of pictures on the next page…"

15.4.2 Representations Externalizing Reasoning

We propose to consider the pictures in Rolfsen's proof as externalizing part of the reasoning: in order to grasp the validity of this proof, it is necessary to imagine the manipulation that leads from one to the other. Moreover, it is the very use of these external representations that triggers *manipulative* imagination, which is, as we will argue, crucial in topology.

This supports an approach to mathematics, according to which mathematical representations and symbols are intimately linked to the concepts they represent. De Cruz and De Smedt (2013, p. 4) claim that "symbols are not merely used to express mathematical concepts" but are "constitutive of the concepts themselves. Mathematical symbols enable us to perform mathematical operations that we would not be able to do in the mind alone, they are epistemic actions."

In Rolfsen's proof, the sequence of pictures externalized the reasoning, allowing us to "see" the transformations involved. These pictures follow some convention, but their pictorial features are also relevant. For this reason, they cannot be considered as purely symbolic, but can be recognized as hybrid representations presenting symbolic as well as visual elements, which are both to be taken into account by the experts. For example, in the last picture of the proof, the curves are represented as lines on the diagram that are the images of the lines of the previous pictures under the applied transformations, but at the same time, they must be interpreted as codes for gluings.

Not only in topology, but also in different mathematical fields, visual and symbolic elements come in different and often complementary degrees. A notation for which a syntax is explicitly defined can reach a higher degree of abstraction and therefore allow for wider generalization. Nonetheless, most of the times, this happens at the cost of losing a straightforward intuitive interpretation and in some cases the very possibility of exploiting intuition.[14]

A paradigmatic example of this is the use of closed curves to represent syllogistic reasoning in its development from Euler to Shin. Peirce's (as well as Shin's) introduction of new conventions has increased the diagrams' expressive power, but at the expense of the visual clarity and the intuitive interpretation of Euler's original diagrammatic system. The new conventions are more arbitrary and the new representations more confusing (Shin, Lemon and Mumma 2013).

[14]Of course, as Giaquinto suggests, it could be that we develop more sophisticated forms of intuition and imagination allowing us to manipulate also arrays of symbols, or syntactic expressions in general (Giaquinto 2007, Ch. 12). We tend to agree with this claim.

Let us go back to Rolfsen's proof, taking into account this balance between visual and symbolic elements. The topological pictures represent geometric objects that are straightforwardly interpreted as topological objects. On the contrary, in the case of Heegaard diagrams, the representation of a topological object is more codified, so as to allow for a stronger syntactic control. As a notation,[15] Heegaard diagrams have the potential to allow for generalizations: we can code all closed compact 3-manifolds with them. Nonetheless, the price for this is that the interpretation of the representations cannot be driven anymore mainly by intuition. We have to be aware of the conventions introduced in order for the diagram to "make sense".

Two examples will help to better clarify how these degrees of intuitive and conventional elements converge in different mathematical diagrams and how algebraic and symbolic reasoning can interact. (1) Knot diagrams present clear visual elements because they "intuitively" represent geometric objects, but at the same time, they allow for a syntactic control (through local moves specifically defined on them) (De Toffoli and Giardino 2014). (2) Commutative diagrams of homological algebra display a more evident syntactic component: these diagrams no longer describe geometric objects but abstract structures and relations. Nevertheless, their arrangement in space is essential and thus visual features also characterize them and allow us to "manipulate" them effectively (De Toffoli, Diagrams in homological algebra, manuscript).

Furthermore, it is clear that not all representations of the same topological object would have the same degree of effectiveness in giving a material form to the relevant reasoning and thus in promoting inference: specific conditions have to be met. First of all, different presentations are suitable for different purposes. Not only we do observe different degrees of symbolic and visual elements, but for a given mathematical object there can be more or less effective representations.

Let us consider a specific example to clarify: diagrams representing the torus. We can choose among various possibilities. On the one hand, Fig. 15.3c is a classical diagram of a torus (in this specific picture two curves are added to it), where just a few lines are easily interpreted as a three-dimensional object. A more detailed picture, for example, depicting thickness or shadows, would make the representation more similar to the corresponding material object, but would be less useful since it would distract the viewer from the essential topological features of the object by adding "noise."[16] The connection between topological and material objects is crucial in topology, but only in so far as it stimulates topological imagination that takes inspiration from the one used to manipulate concrete objects but afterwards develops independently. The similarity with concrete objects has its limits: we must be able to detach ourselves from this analogy and perform an abstraction in order to extract relevant topological features. On the other hand, Fig. 15.3a is definitely

[15]Two features of mathematical diagrams allow for an interpretation of them as a notational system: (i) they follow certain conventions and are a codified way to present different mathematical concepts and (ii) they can be used in sequences and constitute a system through which it can be possible to "calculate" effectively.

[16]See Sullivan (2013) for a survey of different mathematical, in particular topological, pictures.

more complicated for novices (e.g., one must know what the arrow signifies to correctly interpret the diagram), but nevertheless it can be useful. As Colyvan (2012, Ch. 8) claims, this notation presents both geometric and algebraic features (which can be exploited in different ways). A side remark is that the consideration of such representations leads to a rejection of a sharp distinction between algebraic and geometric reasoning in mathematics (Giaquinto 2007, Ch. 12). The abstract character of this hybrid notation opens the way to new discoveries; in fact, it leads to generalizations: not only does the same notation allow us to present the cylinder, the Möbius band, the torus, and the Klein bottle, but if we generalize the notation and instead of a square we take a polygon with n sides, then we can present every closed compact surface (see, for instance, Massey 1991, Ch. 1). Moreover, Colyvan points out at the construction of the Klein bottle. The Klein bottle is straightforwardly presented with the square diagrams (we just have to invert one arrow in the diagram of the torus) and in a sense it is the very notation that drives us toward the study of a surprising new object that is two dimensional, but cannot be embedded in three-dimensional space and does not have an outside or inside.

This example illustrates how for topological diagrams, by adding more syntactic elements we often get a more powerful notation from an expressive point of view and at the same time we lose the analogy with material objects.[17] This is not to say that some diagrams are better than others in principle: it depends on the specific context and on the particular purpose for which they are used. The possibility of choosing among a wide variety of representations enriches the set of tools available to the mathematician.

A final remark about representations concerns their materiality. To avoid confusion, we have to keep in mind the distinction between the material pictures and the imagination process, which, especially in the case of trained practitioners, tends to vanish. Actual topological pictures trigger imagination and help us see modifications on them, but experts may not find it necessary to actually draw all the pictures. The same holds for algebra, where experts skip transitions that non-trained practitioners cannot avoid writing down explicitly. This does not mean that experts do not need pictures to grasp the reasoning, but only that, thanks to training and thus to their familiarity with drawing and manipulating pictures, they are sometimes able to determine what these pictures would look like even without actually drawing them. More generally, for each subfield it would be possible to define a set of "background pictures" that are common to all practitioners. This set of pictures determines what Thurston calls a "mental model" for a group of mathematicians in a subdiscipline (Thurston 1994, pp. 174–175). For instance, any knot theorist knows without need for material pictures what a diagram of the trefoil knot or of the figure-eight knot looks like. To give a more sophisticated example, the original proof by Alexander of his famous theorem about the possibility of representing any knot as a braid is a visual argument that requires the use of this type of imagination but does not contain a single picture (Alexander 1923).

[17] Of course, these syntactic elements are not arbitrary introduced, but according to specific aims.

15.4.3 "Seeing" in Topology Through Manipulative Imagination

We have just claimed that in order to understand Rolfsen's proof, we need to "see" topological transformations. As the text says, "The isotopic transformation (1) – (4) *shows* that this outer part is also a handlebody." To "see" what is "shown" by the picture is crucial. Even in simpler cases, instructions are useful in making us grasp the transformations, as for example going from Fig. 15.9a to Fig. 15.9b.

Therefore, in low-dimensional topology, following an argument, let alone being able to construct one, often requires "seeing" certain transformations. This method of proving, different from the standard one in other areas of mathematics, which is entirely propositional, consists in providing pictures plus instructions for their interpretation. Yet, what does "seeing" actually mean in this context? We can connect it with intuition. Heinzmann, writing about Poincaré's *Analysis Situs*, traces the need for three distinct types of intuition in what is now topology:

> Defining it as the science of classification of closed surfaces, called later manifolds, with respect to continuous deformations, it requires *geometric intuition* concerning the qualitative property of a n-dimensional manifold, *arithmetic intuition* insofar as he introduced computing with the topological object 'manifold' and, insofar as the strongest classification-criterion is the fundamental group, one needs *algebraic intuition*, too. (Heinzmann 1999, p. 55, *emphasis added*)

Even if the term "intuition" is vague, we can specify what it means in the present context: for low-dimensional topology, it is possible to point at some of the "intuitive" capacities involved. Our proposal is that topologists use a special kind of imagination that does not only involve vision but also spatial-motoric intuition of three-dimensional space. In fact, in Rolfsen's proof, one finds no difficult calculations—at least in the standard meaning of the term—and nonetheless the argument is not easy to follow. In order to understand the proof and check its validity, practitioners have to use their ability to imagine topological transformations *correctly*. For example, they have to interpret the transitions between the pictures as homeomorphisms. As a result, "seeing" in topology means first to interpret the representations coherently with the shared practice and then performing epistemic actions on them. In the case of Rolfsen's proof, these actions take the form of continuous deformations. The interaction with the representations is thus pivotal: mathematicians have to activate this form of imagination in order to use the representations as inferential tools.

Consequently, "seeing" is here to be intended as much more than simple vision for at least two reasons. First, because it exploits some of our spontaneous cognitive abilities such as vision, but has nevertheless to be properly trained inside the practice in order to be correctly applied. Secondly, because mathematicians do not only see particular representations, but also the possible actions that could be performed on them, i.e. the possible transitions between pictures. We chose to label this capacity

manipulative imagination[18] as a sophisticated form of imagination that derives from our preexisting manipulative capacities with concrete objects and our motor agency in three-dimensional space: it seems to have a spatial-motoric and not a specifically visual nature.[19] An important feature of this form of imagination is that it is not exclusively innate or *a priori*; on the contrary, it needs to be specifically trained by experimenting with the available representations. This is in line with Poincaré's view that "intuition" has to be trained:

> The main goal in teaching mathematics is to develop some faculties of the mind, and among these, intuition is not the least precious.[20] (Poincaré 1889, p. 160)

Moreover, it is crucial to highlight that it is the massive use of this form of imagination that is responsible for the peculiar development of low-dimensional topology, which has been so different from that of other mathematical fields. As Jones (1998, p. 212) writes, in low-dimensional topology, we do not need to formalize every argument for the very reason that we can rely on our intuition. In other fields, this intuition is unavailable, just as it would be unavailable in low-dimensional topology if we were two-dimensional creatures without the imagination of three-dimensional space. In this case, we would have to formalize each argument and low-dimensional topology would be more similar to more abstract areas of mathematics.

15.4.4 Justifications and Criteria of Validity

As we have already discussed, the practitioners of low-dimensional topology "see" the transformations and check whether they are permissible: the representations embody their reasoning and provide at the same time evidence for their conclusions.

Of course, as in any proof, not everything has to be justified. The background knowledge, amounting to the mental models shared by the members of the community to which the proof is addressed, is assumed as already established. Moreover, particular standards of justification and criteria of validity are provided: the permissible actions on the representations are already defined, and they are part of the background material.

In Rolfsen's proof, we saw that among the permissible actions on the pictures are continuous transformations. These are part of the background material in the sense that any topologist knows immediately that these transformations can be interpreted

[18]We already used the expression "manipulative imagination" when studying knot diagrams (De Toffoli and Giardino 2014).

[19]Think of the blind mathematician Morin who contributed to the understanding of one of the first actual sphere eversions (Morin and Petit 1980).

[20]We have translated from the original: "Le but principal de l'enseignement mathématique est de développer certaines facultés de l'esprit, et parmi elles l'intuition n'est pas la moins précieuse."

in terms of homeomorphisms. The validity is thus based on the "practice": it is the practice itself that integrates a way of controlling the actions on the representations used, which results in the establishment of local criteria for validity. The responsibility is shared among experts: since in low-dimensional topology different forms of reasoning are employed, some of which are specific to it, purely external criteria of validity cannot exhaust all the criteria actually adopted. As Brown suggests, we should acknowledge the existence of non-formal reasoning in mathematics: "first-order logic may be well understood, but what passes for acceptable proof in mathematics includes much more than that" (Brown 1999, p. 164). If this is true, then, as Larvor has exhaustively discussed, "the cost is that we have to abandon the hope of establishing a general test for validity" (Larvor 2012, p. 723).

In our view, what Thurston refers to as "good human processes for checking mathematical validity" (Thurston 1994, p. 171), are context-dependent processes that in low-dimensional topology rely on our manipulative imagination and more generally on our intuition of three-dimensional space, duly trained according to the specific practice. Furthermore, formal proofs are "out of reach," because in order to obtain reliable formalizations, mathematicians would have to spend all of their time to rewrite already known results and conform them to general standards. As Thurston notes, on a small scale, this is easy to do, but on a large scale, where results are interconnected, we would have to check for the coherence of all the arbitrary local choices of formalization. To do so would require a huge amount of time, and topologists are not willing to undertake such a project.

This does not mean that in principle any proof in topology could not be translated so as to assume a propositional and more formal form – even if to do so would be a "nightmare," in Jones' words (Jones 1998, p. 212).[21] The point is that it is not usually done. Nevertheless, as we have mentioned already, part of the confidence of the practitioners is based on the knowledge of how to convey visual modifications in more formal expressions. For example, "gluing" in topology is straightforwardly interpreted in terms of quotient spaces and "deforming continuously" in terms of homeomorphisms. In Thurston's words:

> When people are doing mathematics, the flow of ideas and the social standard of validity is much more reliable then formal documents. People are usually not very good in checking *formal correctness* of proofs, but they are quite good at detecting potential weaknesses or flaws in proofs. (Thurston 1994, p. 169)

We would also like to mention that once we accept that the argumentative form is based on such externalizations of reasoning, given by the specific representations used and the control of the practice, then the process of discovery and that of justification seem to occur at the same time.

Let us now turn to the issue we addressed at the end of the previous section, that is, to the reasons why low-dimensional topology is such a special subfield of mathematics. We have analyzed Rolfsen's proof as a paradigmatic example of

[21] Jones is taking the example of formalizing a proof in knot theory.

an informal argument that can be given in low-dimensional topology, where the sequence of pictures embodies and at the same time justifies the reasoning. Any version of the proof without the pictures, let alone a formal version, would not be able to externalize the reasoning and trigger manipulative imagination and thus would completely obscure the topological permissible actions. By this feature, the case of low-dimensional topology seems to be quite distant from other areas of mathematics. Moreover, this form of reasoning is epistemologically relevant, as Jones makes clear:

> One of the interesting consequences of the use of three-dimensional intuition is that the field of low-dimensional topology has advanced in a way that is significantly different from other branches of mathematics. One is expected to "see" results in this field, and once the result, or partial result, has been "seen," it requires no further discussion. (Jones 1998, p. 212)

At this point, we can interpret Jones' claim in the light of the proposed interpretation. First, manipulative imagination is the cognitive process that modulates our three-dimensional spatial-motoric intuition in relation to the particular mathematical context. Second, thanks to manipulative imagination, we have at our disposal a set of permissible actions.[22] Pictures indicating the stages of transformations, plus instructions explaining how to interpret them, can count as justifications. Exclusively linguistic proofs, and formal proofs are thus just a small portion of the proofs accepted as valid. Of course, it is still possible to translate visual arguments into formal ones. Nevertheless, as Jones and Larvor suggest, the formal version might be complete, but it remains inadequate. As a consequence, once we accept the existence of arguments structured in sequences of pictures, we realize that although there might be good reasons to reduce the reasoning to formal statements, this move would add nothing to the topological reasoning behind the argument.

A practitioner of low-dimensional topology uses material representations, which are the condition for sharing content among the experts. These representations must be adequate to externalize reasoning and to trigger manipulative imagination so as to allow performing permissible and effective inferential actions. To establish the validity of an argument, a low-dimensional topologist shares the responsibility with other practitioners. That is, the community defines criteria of validity specific to their subfield: this is part of the normative structure of this practice.

15.5 Conclusions

In this article, we aimed to show that attention to how low-dimensional topology is practiced gives new insight on the use of mathematical representations. In particular, we unveiled some of the reasons why reasoning with these representations can be seen as an essential part of doing mathematics.

[22]It would be misleading to conceive this set as fixed once and for all, since it varies according to the context of use.

Mathematicians rely on an astonishing variety of proving practices, beyond the one analyzed here. As Larvor suggests (Larvor 2012, p. 723), philosophers should work in the direction of completing the list of all objects involved in mathematical argumentations that are found in the practice. We have also shown that in low-dimensional topology, a proof can take the form of a sequence of pictures accompanied by instructions. The transformations of pictures are the result of permissible epistemic actions. Moreover, the choice of specific representations plays a pivotal role, because it triggers different cognitive skills and externalizes reasoning. In fact, the mathematical practice is characterized by a continuous feedback between specific forms of reasoning and particular representations.

In further research, we aim at comparing different practices from the point of view of the relation between the cognitive abilities triggered, the representations introduced and the argumentative form employed. Another development of the present project will be compiling a taxonomy of topological pictures, which would identify the specific features that are responsible for prompting manipulative imagination through different representational conventions. Moreover, we aim at developing the present framework in order to appreciate the effectiveness of hybrid notations and visual representations, with respect to their possible generalizations and their power to trigger specific kinds of imagination. As Giaquinto claims (Giaquinto 2007, p. 265): "Visual thinking in mathematics is extensive, diverse, familiar, and yet little understood. Here is abundant terrain for research." With the present study, we hope to have given an initial contribution.

Acknowledgements We presented an early version of this work at the first conference on *Mathematical Cultures*, funded by the *Arts and Humanities Research Council* and organized by Brendan Larvor, which took place in London in September 2012. We are thankful to the organizers and to the audience for giving us useful feedback. We thank Ken Manders, John Sullivan and the anonymous referees for their useful comments that helped us improve the article.

References

Adams, C.C., and A.W. Reid. 1993. Quasi-Fuchsian surfaces in hyperbolic knot complements. *Journal of the Australian Mathematical Society* 55: 116–131.

Alexander, J.W. 1923. A lemma on systems of knotted curves. *Proceedings of the National Academy of Science of the United States of America* 9(3): 93–95.

Barwise, J., and J. Etchemendy. 1996. Visual information and valid reasoning. In *Logical reasoning with diagrams*, ed. Allwein and Barwise, 3–25. New York: Oxford University Press.

Brown, J.R. 1999. *Philosophy of mathematics*. London/New York: Routledge.

Colyan, M. 2012. *An introduction to the philosophy of mathematics*. Cambridge: Cambridge University Press.

De Cruz, H., and J. De Smedt. 2013. Mathematical symbols as epistemic actions. *Synthese* 190: 3–19.

De Toffoli, S., and V. Giardino. 2014. Forms and role of diagrams in knot theory. *Erkenntnis* 79: 829–842.

Fomenko, A. 1997. *Algorithmic and computer methods for three-manifolds*. Dordrecht/Boston: Kluwer Academic.

Giaquinto, M. 2007. *Visual thinking in mathematics*. Oxford: Oxford University Press.

Heinzmann, G. 1999. Poincaré on understanding mathematics. In *Philosophia Scientiae, Travaux d'histoire et de philosophie des sciences: Actes du Colloque France-Autriche, Mai 1995: Interférences et transformations dans la philosophie française et autrichienne* (Mach, Poincaré, Duhem, Boltzmann), ed. Heinzmann, 43–60.

Jones, V.F.R. 1998. A credo of sorts. In *Truth in mathematics*, ed. Dales and Oliveri. Oxford: Oxford University Press.

Kirsh, D., and P. Maglio. 1994. On distinguishing epistemic from pragmatic action. *Cognitive Science* 18: 513–549.

Larvor, B. 2012. How to think about informal proofs. *Synthese* 187(2): 715–730.

Mancosu, P. (ed.). 2008. *The philosophy of mathematical practice*. Oxford: Oxford University Press.

Manders, K. 2008. The Euclidean diagram. In *The philosophy of mathematical practice*, ed. Mancosu. Oxford: Oxford University Press.

Massey, W.S. 1991. *A basic course in algebraic topology*. New York: Springer.

Morin, B., and J.P. Petit. 1980. Le retournement de la sphère. In *Les Progrès des Mathèmatiques*, 32–45. Paris/Berlin: Pour la Science.

Poincaré, H. 1889. La logique et l'intuition. *L'enseignement mathématique* 1:157–162, 5.

Poincaré, H. 1900. Second complément à l'Analysis Situs. *Proceedings of the London Mathematical Society* 32: 277–308.

Rolfsen, D. 1976. *Knots and links*. Berkeley: Publish or Perish.

Shin, S-J. 2004. Heterogeneous reasoning and its logic. *The Bulletin of Symbolic Logic* 10(1): 86–106.

Shin, S-J., O. Lemon, and J. Mumma. 2013. Diagrams. In *The Stanford encyclopedia of philosophy*, ed. Zalta, Fall 2013 edition. http://plato.stanford.edu/archives/fall2013/entries/diagrams/.

Sullivan, J.M. 2013. Diagrams and visualization in mathematics. In *Band 4: Sichtbarkeiten 4: Praktiken visuellen Denkens*, ed. von Dieter Mersch, Mira Fliescher, and Fabian Goppelsröder. Berlin: Diaphanes, 15–17 Nov 2012 in Berlin, 2013 forthcoming.

Thurston, W. 1994. On proof and progress in mathematics. *Bulletin of the American Mathematical Society* 30(2):161–177.