22. Diagrammatic Reasoning in Mathematics

Valeria Giardino

The objective of the present chapter will be to review the most recent studies about diagrammatic reasoning in mathematics. Section 22.3 will focus on the very much discussed topic of the role and of the features of diagrams and diagrammatic reasoning in Euclidean geometry. Section 22.4 will be devoted to the proposal of considering diagrams as representations that are introduced in support of other symbolic practices and whose power resides in their ambiguity. In Sect. 22.5, the attention will turn toward studies discussing diagrammatic reasoning in contemporary mathematics. In Sect. 22.6, computational perspectives on how to implement diagrammatic reasoning in computer programs will be introduced, both for Euclidean geometry and theory of numbers. In Sect. 22.7, it will be discussed how the study of diagrammatic reasoning can shed light onto the nature of mathematical thinking in general. Finally, in Sect. 22.8, some brief conclusions about diagrammatic reasoning in mathematics will be drawn. The choice of reviewing the research about diagrammatic reasoning along these lines is of course at least in part arbitrary. The aim of such a regrouping is to provide the reader with a map that can be helpful for exploring the various and already copious literature that has been recently produced on the subject. The ambition is that such a map will be as extensive as possible.

22.1 Diagrams as Cognitive Tools

In his *Parallel Lives*, Plutarch famously reported the murder of Archimedes. He relates three different versions of the circumstances that brought about his death. According to the first one, Archimedes was so intent upon inspecting a diagram to work out some problem that he never noticed the incursion of the Romans, nor that the city was taken. His absorption in study and contemplation of the diagram was so deep that he declined to follow a soldier who had unexpectedly come up to him and commanded him to do so. Given his refusal, the soldier drew his sword and ran him through. In the same spirit, in one of the most celebrated frescoes of the Italian Renaissance, *The School of Athens*, Raphael depicts a group of men attentively watching a scholar – most likely to be interpreted as Archimedes or Euclid – while he draws a geometrical figure on a clay tablet.

The mathematician is thus often portrayed as intensely working on a diagram; this popular image attests to what extent the resource to diagrams, figures, or sketches – among other possible available instruments – is commonly considered as an outstanding element of the practice of mathematics. Is this picture true to the facts? Are diagrams really part and parcel of the mathematical practice? And if it is so, what can be said about their features, use, and relations with other elements of the same practice? The objective of the present chapter is to introduce the most recent works on diagrammatic reasoning in mathematics and to review the answers that have been proposed so far for these questions. In this first section, the domain of inquiry – diagrammatic reasoning in mathematics – and the issues at stake in exploring it will be defined.

First of all, a clarification is needed on the meaning of the term *diagram* in diagrammatic reasoning, so as to avoid misinterpretations. Throughout the chapter – and possibly in contrast with other views – the term will be used in a very broad sense, that is, to include all cases of two-dimensional representations where their two dimensionality is relevant for the way in which information is displayed and read off from them. This seemingly too vague definition is actually appropriate to refer to many different phenomena that are found in mathematics. Moreover, diagrams will be intended here as *cognitive tools* that are meant to spatially display information in order to improve memory and promote inference, and not necessarily to depict mathematical objects. This will have two consequences: first, the focus of the analysis will be on diagrams and not on visualizations; second, lengthy discussions about the implications for the ontology of mathematics will be avoided. For these issues, one can refer among others to *Brown*, who claims that diagrams are not really pictures but rather "windows to Plato's heaven" [22.1, p. 40], or to *Sherry*, who argues that some particular uses of diagrams make a realist view problematic [22.2].

Diagrammatic reasoning is surely relevant for human reasoning in general. As has been pointed out, human reasoning is *heterogeneous*: humans happen to rely on many different sorts of instruments with the aim of externalizing thought, diagrams being among them [22.3]. A common saying is that we are halfway to finding a solution to a problem when we are able to draw the right diagram for it. Nonetheless, in relation to mathematics, it is necessary to distinguish between mere sketches and diagrams. Sketches are certainly widespread and useful for the mathematician to reason about a problem or to communicate with one's peers. However, they will not be the topic of this chapter, which will be devoted to diagrams as parts of a system of representation. Such diagrams obey some (more or less explicit) rules and their manipulation is controlled by the particular practice, in terms that will be defined later.

Not surprisingly, most analyses of diagrammatic reasoning in mathematics have dealt with Euclidean geometry, where the recourse to diagrams is so natural and spontaneous that there is a tendency to take the presence and the effectiveness of diagrams for granted. Moreover, most diagrams in Euclidean geometry become part of our *visual repertoire* from a very early age at school. Think of the *Pythagorean theorem* and the impressive number of so-called *visual proofs* that have been given for it [22.4]. According to this theorem, the square of the hypotenuse (c) of a right triangle equals the sum of the squares of its other two sides (a and b). In letters,

$$a^2 + b^2 = c^2 \, . \tag{22.1}$$

One of the possible visualizations for the Pythagorean theorem is offered in Fig. 22.1.

In Fig. 22.1a, four identical right triangles have been arranged into two rectangles. To obtain a square of side $a + b$, these two rectangles are added to two squares: one of side a, and the other of side b. In Fig. 22.1b, the same four triangles have been rearranged inside the square of side $a + b$ and they now individuate another square of side c. By looking at the two diagrams together and by applying subtraction of the same objects – the four right triangles – to the same object – the square of side $a + b$ – the Pythagorean theorem is obtained.

However, there are cases of diagrammatic reasoning that may be less obvious than in Euclidean geometry, for example, for statements about numerical properties. Consider the following *geometric series*

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = 1 \tag{22.2}$$

and its possible spatial arrangement in Fig. 22.2, in which each new rectangle or square drawn in the diagram – each new element added to the series – brings us closer to the square of area 1 (the example is taken from [22.1, pp. 36–38]).

As Brown points out, this *picture proof* should be contrasted with a traditional proof using ε-δ techniques. In such a proof, we first have to note that an infinite series converges to the sum S whenever the sequence of

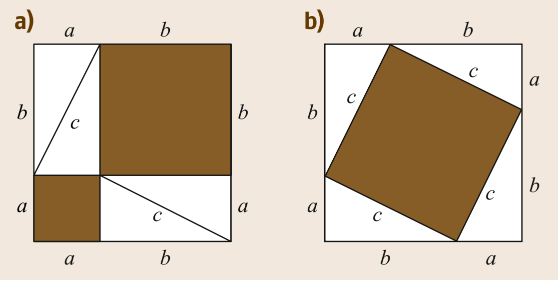

Fig. 22.1a,b Pythagorean theorem

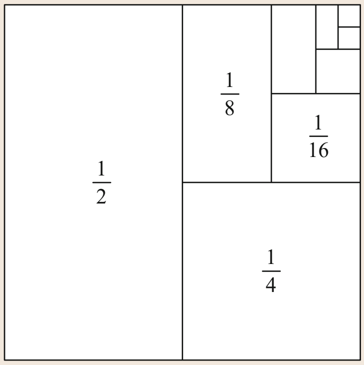

Fig. 22.2 A geometric series

partial sums $\{s_n\}$ converges to S. In this case, we have

$$s_1 = \frac{1}{2}\,, \quad s_2 = \frac{1}{2} + \frac{1}{4}\,, \quad s_3 = \frac{1}{2} + \frac{1}{4} + \frac{1}{8}\,,$$
$$s_n = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots + \frac{1}{2^n}\,. \tag{22.3}$$

The values of these partial sums are

$$\frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \ldots, \frac{2^n - 1}{2^n}\,. \tag{22.4}$$

This infinite sequence has the limit 1, provided that for any number $\varepsilon > 0$, no matter how small, there is a number $N(\varepsilon)$, such that whenever $n > N$, the difference between the general term of the sequence $\frac{2^n-1}{2^n}$ and 1 is less than ε.

In symbols,

$$\lim_{n\to\infty} \frac{2^n - 1}{2^n} = 1$$
$$\iff (\forall\varepsilon)(\exists N)n > N \to \left|\frac{2^n - 1}{2^n} - 1\right| < \varepsilon\,. \tag{22.5}$$

By applying some algebra, one obtains

$$\left|\frac{2^n - 1}{2^n} - 1\right| < \varepsilon \iff \left|\frac{-1}{2^n}\right| < \varepsilon \iff 2^n > \frac{1}{\varepsilon}$$
$$\iff \log_2 \frac{1}{\varepsilon} < n\,. \tag{22.6}$$

Let now $N(\varepsilon) = \log_2 \frac{1}{\varepsilon}$. As a consequence,

$$n > \log_2 \frac{1}{\varepsilon} \to \left|\frac{2^n - 1}{2^n} - 1\right| < \varepsilon\,. \tag{22.7}$$

We have thus proven that the sum of the series is 1. Compare now the easiness of forming the belief that the sum of the series is 1 by looking at the diagram in Fig. 22.2 with the resources required to prove the same result in a traditional way. The topic of the chapter will thus not only be Euclidean geometry. Other studies will be presented that analyze the usefulness of diagrammatic reasoning also in other branches of mathematics.

For the sake of completeness, there exists also very interesting work on ancient mathematics other than in Greece, involving, in some cases, also visual tools [22.5, 6]. Nonetheless, for reasons of space and given the specificity of the research, these works will not be among the subjects of the present chapter. It must also be noted that analogous considerations about the importance of diagrammatic reasoning in mathematics can be made to logic. Many scholars have discussed diagrammatic reasoning in logic, in an interdisciplinary fashion. Some studies have focused on the cognitive impact of diagrams in reasoning [22.7] and others on the importance of heterogeneous reasoning in logical proofs [22.8] and on the characteristics of nonsymbolic, in particular diagrammatic, systems of representation [22.9]. Very recently, and coherently with what will be later said about diagrammatic reasoning in mathematics, it was claimed that different forms of representation in logic are complementary to one another, and that future research should look into more accurate road maps among various kinds of representation so that the appropriate one may be chosen for any given purpose [22.10]. However, for reasons of space and despite the numerous parallels with the case of mathematics, the use of diagrammatic reasoning in logic will not be a topic of the present chapter.

22.2 Diagrams and (the Philosophy of) Mathematical Practice

The subject of diagrammatic reasoning in mathematics has recently gained new attention in the philosophy of mathematics. By contrast, in the nineteenth and twentieth centuries, this topic was neglected and not considered to be of philosophical interest; the heuristic power of diagrams in mathematics was never denied, but visual mathematical tools were commonly relegated to the domain of psychology or to the context of discovery – by referring to a distinction between the context of discovery and that of justification that was very popular and that has become more and more precarious in recent years.

Famously, among others, *Russell* criticized Euclidean geometry for not being rigorous enough from a logical point of view [22.11, p. 404ff]. Consider the very first proposition of the *Elements*, which corresponds to the diagram in Fig. 22.3. The proposition invites the reader to construct an equilateral triangle from a segment AB by tracing two circles with centers A and B, respectively, and then connecting the extremes of the segment with the point that is created at the intersection of the two circles. According to *Russell*, "There is no evidence whatever that the circles which we are told to construct intersect, and if they do not, the whole propositions fails" [22.11, p. 404]. The proposition does, in fact, contain an implicit assumption based on the diagram – the assumption that the circles drawn in the proposition will actually meet. From Russell's and analogous points of view, diagrams do not entirely belong to the formal or logical level, and therefore they should be considered as epistemically fragile. If this is assumed, then a proof is valid only when it is shown to be independent from the corresponding diagram or figure. In order to save Euclidean geometry from the potential fallacies derived from the appeal of diagrams, such as the one just shown, some assumptions, sometimes called *Pasch axioms*, were introduced. For example, it is necessary to assume that *A line touching a triangle and passing inside it touches that triangle at two points*, so as to avoid the reference to the corresponding diagram and make it a logical truth. By contrast, prior to the nineteenth century, such assumptions were generally taken to be "diagrammatically obvious" [22.12, p. 46].

There were historical reasons for this kind of scepticism in relation to the use of visual tools in mathematics. At the end of the nineteenth century, due to progress in disciplines such as analysis and algebra on the one hand, and the development of non-Euclidean geometries on the other, the request for a foundation of mathematics expressed a genuine mathematical need. Euclidean geometry was not the only logically possible geometry, and therefore it did not necessarily convey truth about the physical world: perception, motion, and superposition of figures had to be excluded as illegitimate procedures. In the course of the twentieth century, this *search for certainty* – as *Giaquinto* called it – became a sort of philosophical obsession [22.13]. Figures were considered as definitely unreliable, since they did not any more represent our knowledge of physical space. Moreover, they give rise to errors. Famously, *Klein* presented a case of a diagram that is apparently correct, but which in fact induces one to draw the – false – conclusion that all triangles are isosceles triangles [22.14, p. 202]. Paradigmatic in this sense was *Hilbert's* program, who attempted to rewrite geometry without any unarticulated assumptions [22.15]. For such post-nineteenth century philosophy of mathematics, a proof should be followed, not *seen*.

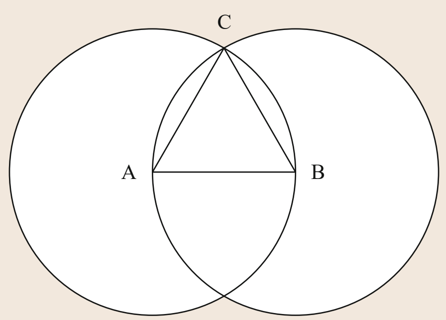

Fig. 22.3 Euclid, Proposition I.1

However, some studies based on the scrutiny of the practice of mathematics have recently challenged this standard point of view. As editors of a book on visualization, explanation, and reasoning styles in mathematics, *Mancosu* et al. explained in 2005 how it was necessary to extend the range of questions to raise about mathematics besides the ones coming from the traditional foundational programs. The focus should be turned toward the consideration of "what mathematicians are actually doing when they produce mathematics" [22.16, p . 1]:

> "Questions concerning concept-formation, understanding, heuristics, changes in style of reasoning, the role of analogies and diagrams etc. have become the subject of intense interest. [...] How are mathematical objects and concepts generated? How does the process tie up with justification? What role do visual images and diagrams play in mathematical activity?"

This invitation to widen the topics of philosophical inquiry about mathematics has developed into a sort of movement, the so-called *philosophy of mathematical practice*, which also criticizes the "single-minded focus on the problem of *access* to mathematical objects that has reduced the epistemology of mathematics to a *torso*" [22.16, p. 1]. Epistemology of mathematics can venture beyond the present confines and address epistemological issues that have to do with [22.16, p. 1]

> "fruitfulness, evidence, visualization, diagrammatic reasoning, understanding, explanation and other aspects of mathematical epistemology which are

orthogonal to the problem of access to *abstract objects*."

This approach would be more in line with what at least some of the very practitioners seem to think about the practice of mathematics. As *Jones*, a topologist and former Field medallist, summarizes, it is quite usual among mathematicians to have very little understanding of its philosophical underpinnings; in his view, for a mathematician, it is actually not at all difficult to live with worries such as Russell's paradox while having complete confidence in one's mathematics [22.17].

In this perspective, the study of diagrammatic reasoning in mathematics thus resumes its philosophical interest, by taking into account the appropriate areas of mathematics. Before presenting the different analyses that have been provided about diagrammatic reasoning in mathematics, three features of diagrammatic reasoning that will characterize most of the studies reviewed should be pointed out. First, diagrammatic reasoning in mathematics is not only *visual* reasoning. In fact, in most cases, a diagram comes with a *text*, and, as a consequence, any analysis of diagrammatic reasoning cannot disregard the role of the text accompanying diagrams. In two very fascinating volumes, *Nelsen* collected a series of proofs, taken from the *Mathematics Magazine*, that he calls "without words" [22.18, 19]. Nonetheless, these proofs are not exactly "without words," since to use a diagram is not only a matter of applying specific perceptual capacities but also of mastering the relevant background knowledge. In Nelsen's proofs, diagrams refer to mathematical statements that can in some way be *found* in them. Diagrams and texts are, in fact, related: each practice will in turn define the terms of this relation. Second, there is another sense in which diagrammatic reasoning is not only visual. In most cases, diagrams are *kinaesthetic* objects, that is, they are intended to be changed and manipulated according to practice. A diagram can be conceived as an experimental ground, where mathematicians are qualified to apply *epistemic actions*, which are – following *Kirsch* and *Maglio's* definition – "actions that are performed to uncover information that is hidden or hard to compute mentally" [22.20]. Third, as will be discussed in the Conclusions, the philosophical interest in studying diagrammatic reasoning is due to the cognitively hybrid status of diagrams. In fact, diagrams are certainly related to text, but at the same time, they are more than a mere visual translation of it; moreover, they are not only synoptic images, but also tools subject to manipulation; finally, they are not only part of the process of discovery, but in the appropriate context of use they are also able to constitute evidence for justification. The inquiry into diagrammatic reasoning in mathematics will in the end force us to blur the standard boundaries between the various elements of the mathematical practice.

22.3 The Euclidean Diagram

A review of the literature on diagrammatic reasoning in mathematics has to start from the research on Euclidean geometry. This section will be thus focused in particular on some of the most influential studies on the role and use of diagrams in the Euclidean system, both from a historical and a cognitive perspective. Given the complexity of such a discussion, details beyond the consideration of diagrammatic reasoning in mathematics will not be treated.

The reason for devoting one whole section of the chapter to Euclidean geometry is that the Euclidean diagram has always been considered as the paradigm of diagrammatic reasoning in mathematics. As *Ferreiros* has proposed, the mathematical practice of Greek geometers summarized in the *Elements* can be considered as a theoretical study of practical geometry [22.21, Chap. 5]. Its theoretical nature comes not only from the new goals and values that are identified as guiding the practice, but also from the idealizations introduced. This picture of Greek geometry contrasts with the abstract tendency of reflections on the subject since Pasch and Hilbert. I have already pointed out in Sect. 22.1 that the post-nineteenth century approach tried to formalize mathematical proofs in such a way that diagrams are not part of them. One of the consequences of such an attitude was to consider diagrams as simple heuristic tools that are possibly useful in illustrating a result, but not constitutive of it. Therefore, there was an interest in translating Euclid's *Elements* – maybe the most widely read text in the entire history of mathematics – into formal sentences of quantificational logic, so as to show that the reference to implicit assumptions based on the diagram could be avoided. A common feature of the studies that will be presented in this section will be precisely to point out that such a move would not represent Euclidean geometry as was originally conceived. If this is true, then it is necessary to provide a plausible explanation for the way in which information that is relevant for the proof can be read off from an Euclidean diagram. The post-nineteenth century philosophy of mathematics gave foundations of logic for what was implicitly assumed in reference to a particular diagram. But what is *implicit* in a diagram? What cognitive abilities are needed to recognize this information and use it in a proof? Some proposals gave a Kantian reading of the *spatial intuition* that is involved in reasoning with a Euclidean diagram, as, for example, in the works of *Shabel* [22.22] and *Norman* [22.23]. According to these views, in Euclid's time, spatial and visual intuition was considered as mathematically reliable, and tacit assumptions were warranted *on the basis of* spatial and visual information. Nonetheless, these works have a wider scope than that of the present chapter, that is, they aim to give evidence in favor of the plausibility of a Kantian philosophy of mathematics, or at least of part of it. For this reason, they will not be discussed here.

Despite the specificity of the Euclidean case, in the remainder of the chapter, it will become evident how some of the characteristics of diagrammatic reasoning in Euclidean geometry can also be adapted to other mathematical practices involving diagrams. As already mentioned, the literature about diagrammatic reasoning in ancient Greek geometry is vast. The studies presented here are among the most influential ones. For other works, one can refer to the bibliography at the end of the chapter and to the references given in the single studies.

22.3.1 The (Greek) Lettered Diagram

The first analysis that will be introduced is the original and fascinating contribution on the shaping of Greek deduction provided by *Netz* [22.12]. Netz' aim is to reconstruct a *cognitive history* of the use of diagrams and text in Greek mathematics. According to his definition, cognitive history lies at the intersection of the history of science and cognitive science: it is analogous to the history of science, because it takes into account cultural artifacts, but it is also comparable to cognitive science because it approaches knowledge not through its specific propositional contents but by looking at its forms and practice. In *Netz*' words, such an intersection is "an interesting but dangerous place to be in" [22.12, p. 7]. In fact, his worry is that historians might see his research as over-theoretical and too open to generalization, while cognitive scientists might consider it as too "impressionistic" [22.12, p. 7].

Netz's idea, in line with the philosophical approach described in Sect. 22.1, is to look at specific practices and consider the influence that they might have (or might have had) on the cognitive possibilities of science. His case study is Greek geometry. Note that Netz' analysis concerns Greek geometry in general and, differently from the studies that will be presented below, does not focus on Euclid only. He starts from the observation that despite the already discussed post-nineteenth century criticisms, when doing Euclidean geometry, one would find it difficult [22.12, p. 23]

> "to *unsee* the diagram, to teach oneself to disregard it and to imagine that the only information there is is that supplied by the text. Visual information is itself compelling in an unobtrusive way."

Euclidean diagrams seem to be part of the visual repertoire of shapes and figures that we are familiar with. If this is the case, then any analysis of Euclidean geometry must take this fact into account. One possible strategy would be to try to reconstruct the geometric practice of the time and focus on what Netz believes is the distinctive mark of Greek mathematics, something that has not been developed independently by any other culture: the *lettered diagram*.

Following Netz' definition, the lettered diagram is a combination of distinct elements that taken together make it possible to generalize an argument that is given in a single diagram having specific geometrical properties. The lettered diagram can thus be considered at different levels. At the logical level, it is composed, as the name suggests, by a combination of the continuous – the diagram – and the discrete – the letters added to it. At the cognitive level, it is a mixture of the visual resources that are triggered by it, and the finite manageable models that the letters made accessible. By following *Peirce's* distinction among icons, indexes, and symbols [22.24], the lettered diagram associates, at the semiotic level, an icon – the diagram – with some indices – the letters. As will be shown in the next sections, Peirce's distinction will be a reference also for other studies on diagrammatic reasoning in mathematics. It is interesting to point out from now that the Peircean terminology, despite being a common background for many of these authors, is applied in a variety of ways to different elements of diagrammatic reasoning in mathematics. The lettered diagram can be considered also from an historical point of view. Against this background, the same diagram is a combination of two elements. First, it refers to an art related to the construction of the diagram which, in Netz' analysis, is most likely a *banausic* art, that is, a practical art serving utilitarian purposes only. Second, it exploits a form of very sophisticated reflexivity, which is related to the use of the letters. The lettered diagram is an effective geometric tool precisely because of the richness of these different aspects characterizing it. In a lettered diagram, we see how almost antagonistic elements are integrated, so as to make it the appropriate instrument to promote and justify deduction [22.12, p. 67].

In Netz' reconstruction, Greek mathematics is constituted by a whole set of procedures for argumentation. These procedures are based on the diagram, which consequently serves as a source of evidence. Thanks to the procedure described in the text accompanying the lettered diagram in Fig. 22.3, one knows that the circles will actually meet at the intersection point. An interesting consequence of this reading is that the lettered diagram supplies a universe of discourse, without referring to any ontological principle. According to Netz, this would be a characteristic feature of Greek mathematics: the proof is done at an object level – the level of the lettered diagram – and no abstract objects corresponding to it need to be assumed. As he explains, in Greek practice [22.12, p. 57]:

> "One went directly to diagrams, did the dirty work, and, when asked what the ontology behind it was, one mumbled something about the weather and went back to work. [...] There is a certain single-mindedness about Greek mathematics, a deliberate choice to do mathematics and nothing else. That this was at all possible is partly explicable through the role of the diagram, which acted, effectively, as a substitute for ontology."

This point on the ontology of the Euclidean diagram is not uncontroversial. Other studies dealing with the Euclidean practice consider it necessary to take into account the abstract objects to which, in a way to define, the diagrams seem to refer. For example, *Azzouni* conjectures that the Greek geometers had to posit an ontology of geometrical objects, even if, in his stipulationalist reading, this drive was not motivated by sensitivity to the presence of anything ontologically independent from us that mathematical terms refer to, but rather by geometers' need to prove things in a greater generality and to make applications easier [22.25]. We will see later how Panza introduces quasi-concrete geometrical objects (Sect. 22.3.4).

In this perspective, the paradox that Netz has to solve is how to explain that one proof – done by referring to a particular diagram, inevitably having specific properties – can be considered as a general result. In his interpretation, a proof in the Greek practice is an event occurring on a papyrus or in a given oral communication, and, despite this singularity, is something that is *felt* to be valid. Nonetheless, validity must be intended here in a different sense than the standard one. When looking at Greek mathematics, and contrary to the post-nineteenth century philosophy of mathematics, logic seems to collapse back into cognition.

In order to reply to this challenge, Netz first points out that generality in Greek mathematics exists only on a *global* plane: a theorem is proved having the global system of Greek mathematics as a background. Thanks to this feature, the proof can be considered as invariant under the variability of the single action of drawing one diagram on the papyrus or of presenting the particular proof orally. Therefore, in Greek mathematics, what counts is the *repeatability* of the proof rather than the *generalizability* of the result (for details, see [22.12, Chap. 5]). According to Netz, to understand Greek geometry, a change of mentality is required: while we are used to generalizing a particular result, Greek mathematicians were used to extending the particular proof to other proofs using other and different objects that are nonetheless characterized by the same invariant elements. A particular construction, given by the lettered diagrams – the diagram plus the text accompanying it – can be repeated, and this is considered as certain.

The lettered diagram was a very powerful tool, because it allowed Greek mathematicians to automatize and elide many of the general cognitive processes that are implied in doing geometry. This was connected to expertise: the more expert a mathematician was, the more immediately he became aware of relations of form and the more readily he read off information from the diagram. Interestingly enough, such a feature of the practice with the Greek diagrams seems to be found in other contemporary mathematical practices as well. As the topologist and former Field medallist *Thurston* has proposed, mathematicians working in the same field and thus familiar with the same practice share the same "mental model" [22.26], which seems to refer precisely to the structure of the particular field and the amount of procedures that can be automatized or elided. To sum up, the diagram is a static object, but it becomes kinaesthetic thanks to the language that refers to it as a constructed and manipulable object: the proof is based on a practical invariance. In Netz' careful analysis, this is the best solution to the problem of generality that could be afforded at the time, given the means of communication at hand. If this is true, then any reconstruction as formalization, such as the one proposed by Hilbert, would not be faithful to the Greek practice. Moreover, Netz argues that Greek mathematics did not deal with philosophical matters. In the sources, nothing like a developed theory supporting this solution can be found.

22.3.2 Exact and Co-Exact Properties

Netz' approach is not the only one based on practical invariances. Consider Manders' contribution in an article that has been – in *Mancosu's* words – "an underground classic" [22.27, p. 14] and that was finally published in 2008 (in its original version, which dates back to 1995) [22.28]. In a later introductory paper, *Manders* presents some of the philosophical issues that emerge from diagrammatic reasoning in geometry [22.29]. For him, Euclidean practice deserves philosophical attention, even only for the simple reason that it has been a stable and fruitful tool of investigation across diverse cultural contexts for over 2000 years. Up to the nineteenth century, no one would have denied that such a practice was rigorous; by contrast, it was rather considered as the most rigorous practice among the various human ways of knowing. Also in Manders' view, the Euclidean practice is based on a distribution of labor between two artifact types – the diagram and the text sequence – that have to be considered together. Note that once again the notion of artifact comes onto the scene as referring to diagrams as well as to text, that is, natural language plus letters linking the text to the diagram. Humans, due to their limited cognitive capabilities, cannot control the production and the interpretation of a diagram so as to avoid any case of alternative responses to it. For this reason, the text is introduced with the aim of tracking equality information. As *Manders* explains, in practice, the diagram and the text share the responsibility of allowing the practitioners to respond to physical artifacts in a "stable and stably shared fashion" [22.28, p. 83].

In Manders' reconstruction, proofs in traditional geometry have two parts: one verbal – the *discursive text* – and the other graphical – the *diagram*. The very objects of traditional geometry seem to arise in the diagram: in his words, "We enter a diagonal in a rectangle, and presto, two new triangles pop up" [22.28, p. 83]. The text ascribes some features to the diagram, and these features are called *diagram attributions*. Letters are introduced to facilitate cross-references between the text and the diagram – also Manders' Euclidean diagram is *lettered*. Defining diagram attributions, Manders introduces a distinction between *co-exact* and *exact* features of the diagram that has become, as will be shown, very influential. A co-exact feature is a directly attributable feature of the diagram, which has certain perceptual cues that are fairly stable across a range of variations. Moreover, such a feature cannot be readily eliminated, thanks to what Manders calls *diagram discipline*, that is, the proper exercise of skill in producing diagrams that is required by the practice. To clarify, if one continuously varies the diagram in Fig. 22.3, its co-exact attributes will not be affected. Imagine deforming the two circles no matter how: this would not change the fact that there still is a point at which the two figures intersect. The distinction thus concerns the control that one can have on the diagram and on its possible continuous deformations. This would be in line with the basic general resource of traditional geometrical practice, that, is diagram discipline: the appearance of diagrams is controlled by standards for their proper production and refinement. Diagram discipline governs the possible constructions.

Consider the features of the diagram of a triangle. Such a diagram would *have to be* a nonempty region bounded by three visible curves, and these curves are straight lines. The first property is co-exact and the second is exact. Paradigmatic co-exact properties are thus features such as a region containing another – unaffected if the boundaries are shifted or deformed – or the existence of an intersection point such as the one required in Euclid I.1, as already discussed. By contrast, exact features are affected by deformation, except in some isolated cases. If one varies the diagram of the equilateral triangle, lines might no longer be straight or angles might lose their equality. In this framework, what is typically alleged as *fallacy of diagram use* rests on reading off from a diagram exact conditions of this kind – for example, that the lines in a triangle are not straight. However, the practice – the diagram discipline – *never allows* such a situation to happen. As already mentioned, practitioners created the resources to control the recourse to diagrams, so as to allow the resolution of disagreement among alternative judgements that are based on the appearance of diagrams, and therefore to limit the risk of disagreement for co-exact attributions. Things become trickier when it comes to exact properties, and this is the reason why the text comes in as support. In fact, since exact attributes are, by definition, unstable under the perturbation of a diagram, they can be priorly licensed by the discursive text. To go back to Euclid I.1, that the curves introduced in the course of the proof are circles is licensed, for example, by Postulate 3; furthermore, it is recorded in the discursive text that other subsequent exact attributions are to be licensed, such as the equality of *radii* (by Definition 15, again in the discursive text).

To sum up, for Manders, the diagram discipline is such that it is able to supervise the use of appropriate diagrams. In the remainder of the chapter, it will be shown how Manders' ideas have influenced other research in diagrammatic reasoning also going beyond traditional Euclidean geometry.

22.3.3 Reasoning in the Diagram

Macbeth has proposed a reading of the Euclidean diagram that is in line with the ones that have just been presented [22.30]. For the purpose of the chapter, it is interesting to note that her aim in reconstructing the practice of Euclidean geometry is to see whether a clarification of the nature of this practice might ultimately tell us something about the nature of mathematical practice in general. She criticizes the interpretation of the

Elements as an axiomatic system and proposes to see it instead as a system of natural deduction. Common notions, postulates, and definitions are not to be intended as premises, but as *rules* or *principles* according to which to reason. Moreover, in her view, a diagram is not an instance of a geometrical figure, but an *icon*. Such a feature of the Euclidean diagram makes the demonstration in the Euclidean system general throughout.

In order to clarify such a claim, Macbeth introduces *Grice's* distinction between *natural* and *nonnatural* meaning [22.31]. For Grice, natural meaning is exemplified by sentences, such as *These spots mean measles*. By contrast, a sentence, such as *Schnee means snow*, expresses nonnatural meaning. Let us suppose then that a drawing is an instance of a geometrical figure, that is a particular geometrical figure. If this is the case, it would have natural meaning and a semantic counterpart. For example, in Fig. 22.3, one sees a particular triangle ABC that is one instance of some sort of geometrical entity called an *equilateral triangle*. But let us instead hypothesize that the drawing has nonnatural meaning and therefore is not an instance of an equilateral triangle but *is taken for* an equilateral triangle. Then, the crucial step would be to recognize the *intention* that is behind the making of the drawing. This is the reason one can also draw an imprecise diagram – for example, drawing a circle that looks like an ovoid – as long as the intention – the one of drawing a circle – is clear. Such an intention is expressed throughout the course of the demonstration. Also, *Azzouni* has suggested that the proof-relevant properties are not the actual (physical) properties of singular diagrammatic figures, but conventionally stipulated ones, the recognition of which is *mechanically executable* [22.25].

To sum up, in Macbeth's reconstruction, the Euclidean diagram has nonnatural meaning and is, by intention, general. Moreover, by following Pierce's distinction again, it is an icon because it *resembles* what it signifies. However, resemblance here cannot be intended as resemblance in appearance. The Euclidean diagram resembles what it signifies by displaying the same relations of parts, that is, by being *isomorphic* to it. The circles in Fig. 22.3 are icons of a geometrical circle because there is a likeness in the relationship of the parts of the drawings. Specifically, the resemblance is in the relation of the points on the drawn circumference to the drawn center compared to the relation of the corresponding parts of the geometrical concept. Such a resemblance can be a feature of the diagram because the geometer means or intends to draw a circle, that is, to represent points on the circumference that are equidistant from the center. Given this intention, it is not important whether or not the figure is precise, that is, whether or not the points on the circumference in the drawn figure really look that way. There is a correspondence between the iconicity of the Euclidean diagram as introduced by Macbeth and co-exact properties in Manders' terms. Also in Macbeth's reading, the diagram is intended to show the relations that are constitutive of the various kinds of geometrical entities involved. As she summarizes, "A Euclidean diagram does not instantiate content but instead formulates it" [22.30, p. 250].

Finally, Macbeth aims to show that the chain of reasoning in Euclidean geometry involving diagrams is not diagram-based but *diagrammatic*. According to her terminology, a reasoning is diagram-based when its moves are licensed or justified by the diagram; by contrast, it is diagrammatic when the mathematician is asked to reason *in* the diagram. Consider again Fig. 22.3. There is a sense in which this figure is analogous to the Wittgensteinian duck–rabbit picture, where one alternates between seeing it as the picture of a duck and seeing it as the picture of a rabbit. In a similar fashion, in order for the demonstration to go through, the mathematician has to alternate between seeing certain lines in the figure as icons of *radii* – and therefore equal in length – and as icons of the sides of a triangle – so as to draw the conclusion that the appropriately constructed triangle is in fact equilateral. The point then is that the physical marks on the page have the potential to be regarded in radically different ways. By pointing at such a feature of the Euclidean diagram, Macbeth aims to make sense of Manders' view, saying that geometrical relations *pop out* of the diagram as lines are added to it. The mathematician uses the diagram to reason *in* it and to make new relations appear.

Moreover, according to Macbeth, the Euclidean diagram has three levels of articulation in the way it can be parsed by the geometer's gaze. At a first level, there are the primitive parts: points, lines, angles, and areas. At the second level, there are geometrical objects that are intended to be represented in the diagram. At the third level, there is the whole diagram, which is not in itself a geometrical figure but, in some sense, contains the objects at the other levels. In the course of the demonstration, the diagram can thus be configured and reconfigured according to different intermediate wholes. Thanks to such a function of diagrams, significant and often surprising geometrical truths can be proved. In Macbeth's account, the site of reasoning is the diagram, and not the accompanying text. Her conclusion is that Euclidean geometry is [22.30, p. 266]

> "a mode of mathematical enquiry, a mathematical practice that uses diagrams to explore the myriad discoverable necessary relationships that obtain among geometrical concepts, from the most obvious to the very subtle."

Another more recent study has complemented Manders and Macbeth's account by emphasizing even more strongly how the Euclidean diagram has a role of *practical synthesis*: to draw a figure means to balance multiple desiderata, making it possible to put together insight – that is timeless – and constructions – that are given in time [22.32]. We also mention here that *Macbeth* has applied similar arguments to the role of Frege's *Begriffschrift* as exhibiting the inferentially articulated contents of mathematical concepts [22.33]. Despite the interest of this account, for the reasons given in Sect. 22.1, we will not give here the details of such a study.

In Sect. 22.4, we will come back to the notion of iconicity and see how the productive ambiguity to which Macbeth alludes to in talking about the parsing of the Euclidean diagram can also be found in other cases of diagrammatic reasoning.

22.3.4 Concrete Diagrams and Quasi-Concrete Geometrical Objects

Another view on the generality of the Euclidean diagrams has recently been proposed by *Panza* [22.34]. His aim is to analyze the role of diagrams in Euclid's plane geometry, that is, the geometry as expounded by Euclid in the first six books of the *Elements* and in the *Data*, and as largely practiced up to early-modern age (see also [22.35]). In his view, Euclid's propositions are general insofar as they assert that there are some admitted rules that have to be followed in constructing geometric objects. Once again, what matters for generality are construction procedures. These admitted rules allow the geometer to construct an object having certain properties and relations. To put it briefly, it would be impossible for one to follow the rules and end up with constructing an object without the requested properties.

Panza argues that arguments in the Euclidean system are *about* geometrical objects: points, segments of straight lines, circles, plane angles, and polygons. Taking inspiration from *Parsons* [22.36], he defines such geometrical objects as *quasi-concrete*. Their quasi-concreteness depends precisely on the relation they have with the relevant diagrams, which are instead *concrete* objects: the Euclidean diagram is a configuration of points and lines, or better is what is common to equivalence classes of such configurations. Two claims describe the peculiarity of the relation between quasi-concrete geometrical objects and concrete diagrams. First, the identity conditions of the geometrical objects are provided by the identity conditions of the diagrams that represent them. In his definition, this is the *global* role of diagrams in Euclid's arguments: a diagram is taken as a starting point of licensed procedures for drawing diagrams and a geometrical object can be given in the Euclidean system when a procedure is stipulated for drawing a diagram representing it. Second, the geometrical objects inherit some properties and relations from these diagrams. This is the *local* role of Euclidean diagrams. Such properties and relations are recognized because a diagram is compositional. So understood, a diagram is a configuration of concrete lines drawn on an appropriate flat material support. According to Panza, Euclid's geometry is, therefore, neither an empirical theory nor a contentual one in Hilbert's sense, that is, a theory of "extra-logical discrete objects, which exist intuitively as immediate experience before all thought" [22.37, p. 202]. In his view, differently from the approaches described so far, it is crucial to define an appropriate ontology for the Euclidean diagram. In fact, his objective is to argue against the view that arguments in Euclid's geometry are not about singular objects, but rather about something like general schemas, or only about concepts. Such a view, according to which Euclidean geometry would deal with purely ideal objects, is often taken to be Platonic in spirit and is supposed to have been suggested by *Proclus* [22.38, 39]. Panza's proposal is instead closer to an Aristotelian view that geometric objects result by abstraction from physical ones, but the author claims that it is not his intention to argue that Euclid was actually guided by an Aristotelian rather than a Platonic insight.

In the same spirit, also *Ferreiros* suggests that the objects of Greek geometry are taken to be the diagrams and other similarly shaped objects [22.21, Chap. 5]. Of course, the diagram in this context is not intended to refer to the physically drawn lines that are empirically given, but to the *interpreted* diagram, which is perceived by taking into account the idealizations and the exact conditions conveyed in the text and derived from the theoretical framework in the background. For the geometer, the figure one works with is not intended as an empirical token but as an ideal type. Nonetheless, it is crucial to remark that such an ideal type does not exist outside the mind of the geometer and becomes available only thanks the diagram. Therefore, on the one hand, the object of geometry is the diagram, and, as a consequence, the diagram *constitutes* the object of geometry; on the other hand, the diagram has to be interpreted in order to make the object emerge, and accordingly it also *represents* the object of geometry. Moreover, quoting Aristotle, Ferreiros points out that Greek geometry remains a form of theoretical and not practical geometry, for the reason that its objects are conceived as *immovable and separable*, without this necessarily leading to the thesis that there exist *immovable and separable* entities [22.40].

22.4 The Productive Ambiguity of Diagrams

This brief section will be devoted to the discussion of the role of ambiguity in diagrammatic reasoning. Grosholz has devoted her work to develop a pragmatic approach to mathematical representations, by arguing that the appropriate epistemology for mathematics has to take into account the pragmatic as well as the syntactic and semantic features of the tools that are used in the practice of mathematics. The post-nineteenth century philosophy of mathematics wants all mathematics to be reduced to logic; by contrast, Grosholz claims that philosophy should account for all kinds of mathematical representations, since they are all means to convey mathematical information. Moreover, the powers and limits of each of them should be explored. One format might be chosen among the others for reasons of convenience, depending on the problem to solve in the context of a specific theory or in a particular historical moment. Even the analysis of the use of formal language can thus be framed in terms of its representational role in a historical context of problem-solving. As *Grosholz* explains [22.41, p. 258],

> "Different modes of representation in mathematics bring out different aspects of the items they aim to explain and precipitate with differing degrees of success and accuracy."

In such a picture, a central cognitive role is played in mathematics by a form of controlled and highly structured ambiguity that potentially involves all representations, and is particularly interesting in the case of diagrams. Grosholz as well adopts the general Peircean terminology and distinguishes between *iconic* and *symbolic* uses of the same representation. These two different uses make representations potentially ambiguous.

To clarify, consider as an example Galileo's treatment of free fall and projectile motion in the third and fourth days of his *Discourses and Mathematical Demonstrations Concerning Two New Sciences*. Galileo draws a geometrical figure to prove that (Fig. 22.4),

> "The spaces described by a body falling from rest with uniformly accelerated motion are to each other as the squares of the time-intervals employed in traversing these distances."

In the right-hand figure, the line HI stands for the spatial trajectory of the falling body, but is articulated into a sort of ruler, where the intervals representing distances traversed during equal stretches of time, HL, LM, MN, etc., are indicated in terms of unit intervals, which are represented by a short cross-bar, and in terms of intervals, whose lengths form the sequence of odd numbers $(1, 3, 5, 7, \ldots)$, which are represented by a slightly longer cross-bar. The unit intervals are intended to be counted as well as measured. In the left-hand figure, AB represents time, divided into equal intervals AD, DE, EF, etc., with perpendicular instantaneous velocities raised upon it – EP, for example, represents the greatest velocity attained by the falling body in the time interval AE – generating a series of areas which are also a series of similar triangles. Thanks to an already proven result from Th. I, Prop.1, Galileo builds the first proposition, according to which the distance covered in time AD (or AE) is equal to the distance covered at speed $1/2$ DO (or $1/2$ EP) in time AD (or AE). Therefore, the two spaces that we are looking for are to each other as the distance covered at speed $1/2$ DO in time AD and the distance covered at speed $1/2$ EP in AE. Th. IV, Prop. IV tells us that "the spaces traversed by two particles in uniform motion bear to one another a ratio which is equal to the product of the ratio of the velocities by the ratio of the times"; in this case, given the similarity of the triangles ADO and AEP, AD and AE are to each other $1/2$ DO and $1/2$ EP. Then, the proportion between the two velocities compounded with the time intervals is equal to the proportion of the time intervals compounded with the time intervals, and therefore $[V_1 : V_2]$ compounded with $[T_1 : T_2]$ equals $[T_1 : T_2]^2$. As a consequence, the spaces described by the falling body are proportional to the squares of the time intervals: $[D_1 : D_2]^2 = [T_1 : T_2]^2$. Look now at the left-hand diagram. Consider the sums

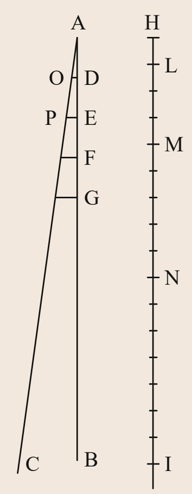

Fig. 22.4 Galileo, *Discorsi*, third day, naturally accelerated motion, Theorem II, Proposition II

$1+3=2^2$, $1+3+5=3^2$, $1+3+5+7=4^2$, and so forth. These sums represent distances and are proportional to the squares of the intervals. Therefore, the time elapsed is proportional to the final velocity and the distance fallen will be proportional to the square of the final velocity.

Galileo's use of the diagram can be analyzed in relation to the different *modes of representation* that are employed to express his argument to prove the theorem. First, he refers to at least four modes of representation: proportions, geometrical figures, numbers, and natural language. Second, the same geometrical diagram serves as an icon and at the same time as a symbol. As an icon, it is configured in such a way that it can stand for a geometrical figure and exhibit patterns of relations among the data it contains. For example, when proportions are taken as finite, they are represented iconically. When the proportions are taken as infinitesimal (because one may take "any equal interval of time whatsoever" [22.41, p. 14]), the diagram is instead used as a symbol. In this case, the configuration of the diagram changes because it is now intended to represent dynamical, temporal processes. Therefore, despite the fact that an appropriate parsing of the diagram cannot represent iconically something that is dynamical or temporal, it can still do it symbolically. In Grosholz' view, the distinction between iconic and symbolic use of a mode of representation sheds light on the importance of semantics in mathematics. In fact, for a mode of representation to be intended not only iconically but also symbolically, the reference to some background knowledge is necessary. The representation does not have to be intended in its *literal* configuration but from within a more elaborated context of use, which provides a new interpretation and a new meaning for it. Galileo's diagram must thus be interpreted in two ways: intervals have to be seen as finite – so that Euclidean results can be applied – and also as infinitesimals – so as to represent accelerated motion. In the proof, errors are prevented by a careful use of ratios.

Compare this example with Macbeth's discussion of Euclid I.1 in the previous section. Here as well, there is only one set of diagrams, but, in order for the demonstration to go through, it must be read and interpreted in different ways. However, Macbeth and Grosholz employ Peirce's distinction in a different way. Macbeth talks of Euclidean diagrams as icons for geometrical relations, while Grosholz refers to two possible different uses – iconic or symbolic – of the same diagram. Moreover, Macbeth's Gricean distinction between natural and nonnatural meaning does not coincide with the distinction between the literal and nonliteral – conventional – uses of the representation made here by Grosholz.

Grosholz' approach is not limited to diagrams in mathematics, unless one wants to say that all mathematical representations are diagrammatic. In fact, in her view, another straightforward example of productive ambiguity is Gödel's representation of well-formed formulas through natural numbers, whose efficacy stems from their unique prime decomposition. In her terminology, the peculiarity of Gödel's proof of incompleteness is that the numbers in it must stand iconically for themselves – so as to allow the application of number theoretic results – and symbolically for well-formed formulas – so as to allow transferring those results to the study of completeness and incompleteness of logical systems. Without going into details, it is sufficient to say that Grosholz points out that this particular case shows how much even logicians exploit the constitutive ambiguity of some of the representations they use. In her view, the recourse to ambiguous formats is, in fact, typical of mathematical reasoning in general, and this is precisely the feature of mathematics that has not been recognized by the standard post-nineteenth century approaches, which have focused on the possibility of providing a formal language that would avoid ambiguities. As *Grosholz* explains [22.41, p. 19]

> "the symbolic language of logistics is allegedly an ideal mode of representation that makes all content explicit; it stands in isomorphic relation to the objects it describes, and that one-one correspondence insures that its definitions are 'neither ambiguous nor empty'."

In Grosholz's view, ambiguity and iconicity then seem to be not only a mark of diagrams such as Galileo's one, but also crucial features of mathematical representations – formulas not being an exception.

In the following sections, other examples of productive ambiguity and iconicity in contemporary mathematics will be given.

22.5 Diagrams in Contemporary Mathematics

As shown in the previous sections, most examples of diagrams that have been discussed in the literature so far are taken from the history of mathematics; furthermore, the focus has been on geometric diagrams. It is worth mentioning here an interesting study by *Chemla* about Carnot's ideas on how to reach generality in geometry, where she analyzes Carnot's treatment of the so-called *theorem of Menelaus* [22.42]. In her reconstruction, Carnot believes that, at least in the case of the theorem of Menelaus, the diagram must be considered as a configuration, appropriately chosen with the aim of finding the solution to the problem in question. As a consequence, the theorem no longer concerns a specific quadrilateral, but any intersection between a triangle and a straight line. Chemla claims that Carnot's ideas were nonstandard at his time, because he introduced a way of processing information that relies on individuating what a general diagram is in opposition to a multitude of particular figures.

This section will be devoted to briefly introducing some works on diagrammatic reasoning in present-day mathematics. The studies have been divided into three categories: analysis, algebra, and topology. Differently from the Euclidean or the theory of number case, the examples taken from contemporary mathematics deserve much more technical machinery in order to be understood, that is, even only to introduce the diagram, much mathematics is required. For reasons of space, it is therefore impossible to give here all the mathematical details, and I invite the reader to refer to the original papers.

22.5.1 Analysis

In two different articles, *Carter* analyzed a case study of diagrammatic reasoning in free probability theory, an area introduced by Voiculescu during the 1980s [22.43, 44]. The aim of free probability theory was to formulate a noncommutative analog to classical probability theory, with the hope that this would lead to new results in analysis. In particular, Carter discusses a section of a paper written by *Haagerup* and *Thorbjørnsen* [22.45], where a combinatorial expression for the expectation of the trace of the product of so-called *Gaussian random matrices* (GRMs) of the following form is found

$$\mathrm{E} \circ Tr_n[B^* B^p] . \tag{22.8}$$

The authors show that this expression depends on the following

$$\mathrm{E} \circ Tr_n[B_1^* B_{\pi(1)} \dots B_p^* B_{\pi(p)}] . \tag{22.9}$$

The indices $\pi(i)$ are symbols denoting the values of a permutation π on $\{1, 2, \dots, p\}$. Therefore, the value of the expression depends on the existence and properties of the permutation that pairs the matrices off 2×2.

Diagrams can be introduced to represent the permutations, and this is a crucial move, since such diagrams make it possible to study permutations independently from the fact that they were set forth as indices of the GRM. Moreover, the recourse to diagrams makes it easier to evaluate the properties of the permutations. Once the relevant properties of the permutation are identified, thanks to the diagram, they can then be reintroduced into the original setting.

To give an idea of what the diagrams representing permutations look like, consider two examples of constructing the permutation $\hat{\pi}$. Let $p = 4$, so that $\pi : \{1, 2, 3, 4\} \to \{1, 2, 3, 4\}$.

Instead of writing

$$B_1^* B_{\pi(1)} B_2^* B_{\pi(2)} B_3^* B_{\pi(3)} B_4^* B_{\pi(4)} , \tag{22.10}$$

we rewrite the expression in the following form

$$C_1^* \cdot C_2 \cdot C_3^* \cdot C_4 \cdot C_5^* \cdot C_6 \cdot C_7^* \cdot C_8 . \tag{22.11}$$

Suppose then that $\pi(1) = 2$ and $\pi(3) = 4$, giving

$$B_1^* \cdot B_2 \cdot B_2^* \cdot B_1 \cdot B_3^* \cdot B_4 \cdot B_4^* \cdot B_3 . \tag{22.12}$$

What the permutation $\hat{\pi}$ is supposed to do is to tell us which of the Cs are identical, in terms of their indices. By comparing the two expressions, we see that $C_1 = C_4$, $C_2 = C_3$, $C_5 = C_8$, and $C_6 = C_7$. In terms of the permutation $\hat{\pi}$, this means that $\hat{\pi}(1) = 4$ and $\hat{\pi}(2) = 3$, and so on. Both permutations can be represented by the diagrams in Fig. 22.5.

Another example could be $\pi(1) = 3$ and $\pi(2) = 4$, giving

$$B_1^* \cdot B_3 \cdot B_2^* \cdot B_4 \cdot B_3^* \cdot B_1 \cdot B_4^* \cdot B_2 . \tag{22.13}$$

By rewriting it in terms of C_i's and comparing again, we obtain $C_1 = C_6$, $C_2 = C_5$, $C_3 = C_8$, and $C_4 = C_7$, as shown in Fig. 22.6.

First, diagrams would suggest definitions and proof strategies. In Carter's example, the definitions of a pair of neighbors, or of a noncrossing and a crossing permutation as well as of cancellation of pairs – manipulations that are all clearly visible in the diagrams – are *inspired* by them. Moreover, as confirmed by the very authors of the study, also the formal version of at least a part of the proofs is inspired by the proof based on

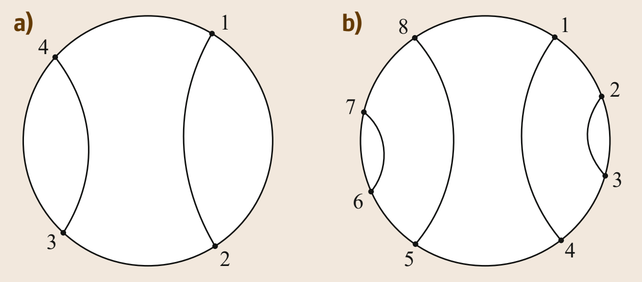

Fig. 22.5 **(a)** π is the permutation (12)(34); **(b)** the correspondent $\hat{\pi}$

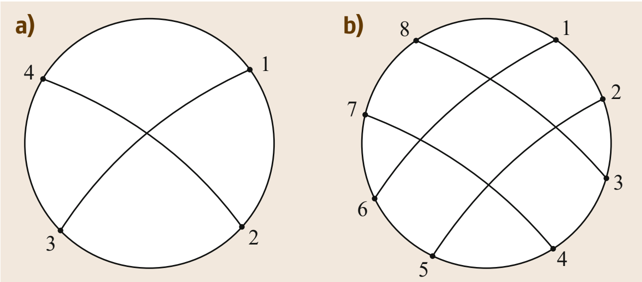

Fig. 22.6 **(a)** π is the permutation (13)(24); **(b)** the correspondent $\hat{\pi}$

diagrams. Second, diagrams function as *frameworks* in parts of proofs: Although they are not used directly to give rigorous proofs, they still play an essential role in the discovery and formulation of both mathematical theorems and proofs, and thus in the practice of the mathematical reasoning.

Carter's idea is that certain properties of the diagrams correspond to formal definitions. In her case study, some diagrams are used to represent permutations and similar diagrams to represent equivalence classes. Diagrams thus make it possible to perform *experiments* on them; for example, the crossings identify the number of the equivalence classes and therefore the definition of a crossing is given an algebraic formulation. Likewise, the concepts of a neighboring pair and of removing pairs (from the diagram) are translated into an algebraic setting. To sum up, the relations used in the proof based on the diagram represent relations that *also* hold in the algebraic setting. As Carter explains, the notions of crossing and neighboring pairs are, in Manders-inspired terminology, examples of coexact properties of the diagrams. In a Piercean semiotic perspective, the diagram in this case would again be *iconic*, and it is for this reason that one can translate the diagrammatic proof into an algebraic proof. In this example, from contemporary mathematics, we are in a sense certainly far from the Euclidean diagram, but we still see that the proof includes an accompanying text; only when the appropriate text is added, do text and diagram – taken together – constitute a proof. The text is also important to disambiguate diagrams that can be interpreted as representing different things (recall Manders' view on the Euclidean diagram).

In a more recent article, Carter discusses at length her reference to Peirce's terminology. Her reconstruction of Peirce's discussion of the use of representation in mathematics is based on some of the most recent studies about Peirce's mathematical philosophy [22.46]. Note that the central notion for *Peirce* is the one of *sign*, that is, in his words, "Something that stands for something else" [22.24, 2.228]. A sign can stand for something else not in virtue of some of its particular features, but thanks to an interpretant that links the sign to the object. For Peirce, signs are then divided into three categories: icons, indices, and symbols; icons are signs in virtue of a relation of likeness with their objects, indices are actually connected to the objects they represent, and symbols represent an object because of a rule stipulating such a relation. Central to Peirce's conception of reasoning in mathematics is that all such reasoning is *diagrammatic* – and therefore iconic. Moreover, Peirce employs the term *diagram* in a much wider sense than usual. In his view, even spoken language can be diagrammatic. Consider a mathematical theorem that contains certain hypotheses. By fixing the reference with certain indices, it is possible to produce a diagram that displays the relations of these referents. In statements concerning basic geometry, the diagram could be a geometric diagram such as the Euclidean diagram. But in other parts of mathematics, it may take a different form. In Carter's view, the diagrams in her case study are iconic because they display properties that can be used to formulate their algebraic analogs. Moreover, the role of indices – the numbers – in the diagram is to allow for reinserting the result into its original setting. Once such a framework is assumed, then diagrams as well as other kinds of representations used in mathematics become an interesting domain of research. As already discussed when presenting Grosholz's work, the objects of inquiry extend from mathematical diagrams to mathematical signs – mathematical representations – in general, including, for example, also linear or two-dimensional notations. In the final section, we will say more about this issue. A further point made by Carter is that the introduced diagrams also enable us to break down proofs into manageable parts, and thus to focus on certain details of a proof. By using diagrams at a particular step of the proof, one needs only to focus on one component, thus getting rid of irrelevant information. In an unpublished paper, *Manders* makes a similar point by introducing the notions of responsiveness and indifference in order to address the topic of progress in mathematics [22.47]. In the following section, more details about this paper will be given.

It is interesting to note that Carter discusses a potential ambiguity of the term *visualization*, used as (i) representation, as in the example given, and as (ii) mental picture, helping the mathematician see that something is the case. In this second meaning, diagrams would be *fruitful frameworks* to trigger imagination. Carter's claim is that there is not a sharp distinction to be drawn here between concrete pictures and mental ones, but quite the opposite: a material picture may trigger our imagination, producing a mental picture, and vice versa a mental picture may be reproduced by a concrete drawing. We will come back also to this issue later.

22.5.2 Algebra

Another case study from contemporary mathematics is taken from a relatively recent mathematical subject: *geometric group theory*. *Starikova* has discussed how the representation of groups by using *Cayley graphs* made it possible to discover new geometric properties of groups [22.48, 49]. In this case study, groups are represented as graphs. Thanks to the consideration of the graphs as metric spaces, many geometric properties of groups are revealed. As a result, it is shown that many combinatorial problems can be solved through the application of geometry and topology to the graphs and by their means to groups.

The background behind Starikova's work is the analysis proposed by *Manders* in the unpublished paper already mentioned in presenting Carter's work [22.47]. In this paper, Manders elaborates more on his study on Euclidean diagrams, this time taking into account the contribution of Descartes' *Géométrie* compared to Euclid's plane geometry. He gives particular stress to the introduction of the algebraic notation. In fact, in mathematical reasoning, we often produce and respond to artifacts that can be of different kinds: natural language expressions, Euclidean diagrams, algebraic or logical formulas. In general, mathematical practice can be defined as the control of the *selective responses* to given information, where response is meant to be *emphasizing* some properties of an object while *neglecting* others. According to Manders, artifacts help to implement and control these selective responses, and therefore their analysis is crucial if the target is the practice of the mathematics in question. Moreover, selective responses are often applied from other domains. Think of the introduction of algebraic notation to apply fast algebraic algorithms. In Descartes' geometry, geometric problems are solved through solving algebraic equations, which represent the geometric curves. Also here, the idea is that by using different representations of the same concepts, new properties might become noticeable. Starikova's study would show a case where a change in representation is a valuable means of finding new properties: drawing the graphs for groups would help discovering new features characterizing them. In this perspective, mathematical problem-solving involves the creation of the right strategies of selection: at each stage of practice, some information is taken into account and some other information is disregarded. It is only by responding to some elements coming from the mathematical context and not paying attention to others that we can control each step of our reasoning. Of course, this control and coordination may have different levels of *quality* across practices. Manders' conclusion is that mathematical progress is based on this coordinated and systematic use of responsiveness and indifference, and that such a coordination is implemented by the introduction and the use of the various representations. The role of the accompanying text is still crucial, since diagrams are produced according to the specifications in the text. Thanks to the text, the depicted relations become reproducible and therefore stable; diagram and text keep supporting each other.

To give the reader an idea of what a Cayley graph for a group looks like, we consider first the definition of a *generating set*. Let G be a group. Then, a subset $S \subseteq G$ is called a generating set for the group G if every element of G can be expressed as a product of the elements of S or the inverses of the elements of S. There may be several generating sets for the same group. The largest generating set is the set of all group elements. For example, the subsets $\{1\}$ and $\{2,3\}$ generate the group $(\mathbb{Z},+)$.

A group with a specified set of generators S is called a *generated group* and is designated as (G,S). If a group has a finite set of generators, it is called a *finitely generated group*. For example, the group $\mathbb{Z}$ is a finitely generated group, for it has a finite generating set, for example, $S=\{1\}$. The generated group $\mathbb{Z}$ with respect to the generating set $\{1\}$ is usually designated as $(\mathbb{Z},\{1\})$. The group $(\mathbb{Q},+)$ of rational numbers under addition cannot be finitely generated. Generators provide us with a *compact* representation of finitely generated groups, that is, a finite set of elements, which by the application of the group operation gives us the rest of the group.

We can now define a Cayley graph. Let (G,S) be a finitely generated group. Then the Cayley graph $\Gamma(G,S)$ of a group G with respect to the choice of S is a directed colored graph, where vertices are identified with the elements of G and the directed edges of a color s connect all possible pairs of vertices (x,sx), $x \in G$, $s \in S$.

In the following, we can see three examples of Cayley graphs: the Cayley graph for the first given example, $(\mathbb{Z},\{1\})$, that is, an infinite chain (Fig. 22.7), another Cayley graph for the same group $\mathbb{Z}$ with generators $\{1,2\}$, which can be depicted as an infinite ladder (Fig. 22.8), and finally the Cayley graph for the group $(\mathbb{Z}\{2,3\})$ (Fig. 22.9). By *geometric* properties of groups, Starikova intends the properties of groups that can be revealed by thinking of their corresponding Cayley graphs as metric spaces. In other words, the idea is to look at groups *through* their Cayley graphs and try to see new (geometric) properties of groups, and then

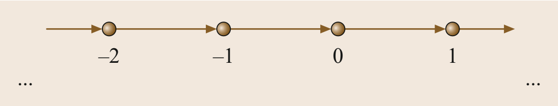

Fig. 22.7 The Cayley graph of the group $(\mathbb{Z}\{1\})$

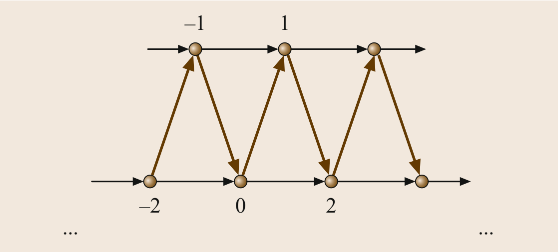

Fig. 22.8 The Cayley graph of the group $(\mathbb{Z}\{1, 2\})$, where bold stands for $\{1\}$

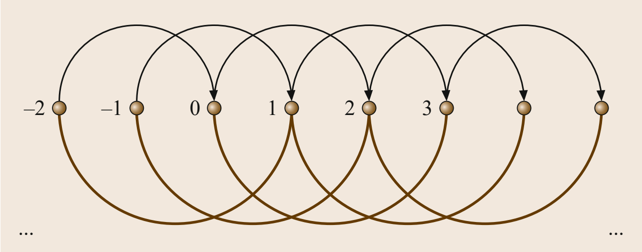

Fig. 22.9 The Cayley graph of the group $(\mathbb{Z}\{2, 3\})$, where bold stands for $\{3\}$

to return to the algebra and check which groups share these properties and under which constraints. Many of these geometric properties turn out, in fact, to be independent from the choice of generators for a Cayley graph. For this reason, they are considered to be the properties of the groups themselves. Such a practice of introducing graphs to represent groups makes it possible to place groups in the same research–object category as classical metric spaces. This can happen because, in Manders's terminology, we can be indifferent to the discrete structure of the group metric space and at the same time respond to the perceptual similarity of particular Cayley graphs having the same metric space. These responses would be unavailable to the combinatorial approach. Moreover, when responding geometrically to Cayley graphs, we perceive them as objects embedded in a space and having geometric elements. But then the response is *modified*, and some diagrammatic features are neglected in order to highlight more abstract properties. By introducing Cayley graphs, a group theorist thus has the opportunity to use them to define a metric of the group and then exploit its geometry, to define geometric counterparts to some algebraic properties of the group, and finally to classify groups having these geometric properties. This case study would show how sometimes the right choice of representation of an abstract object might lead to a significant development of a key concept.

22.5.3 Topology

Other case studies from contemporary mathematics concern topology.

The first one focuses on the identification and the discussion of the role of diagrammatic reasoning in *knot theory*, a branch of topology dealing with knots. A *knot* is a smooth closed simple curve in the Euclidean three-dimensional space, and a *knot diagram* is a regular projection of a knot with relative height information at the intersection points. *De Toffoli* and *Giardino* have discussed how knot diagrams are *privileged* points of view on knots: they display only a certain number of properties by selecting the relevant ones [22.50]. In fact, a single knot diagram cannot exhaust all the information about the knot type, and, for this reason, it is necessary to look at many diagrams of the same knot in order to *see* its different aspects. For example, both diagrams in Fig. 22.10 represent the unknot – that is, as the name suggests, a not knotted knot type – and we can transform the first into the second by *pulling* down the middle arc. However, this move alone does not allow us to conclude that both diagrams represent the unknot; to see that, we would have to apply further similar moves. In the article, a formalization for these possible modifications is provided.

The general idea behind this work is that diagrams are *kinaesthetic*, that is, their use is related to procedures and possible moves imagined on them. In topology, which is informally referred to as *rubber-band geometry*, a practitioner develops the ability to imagine continuous deformations. Manipulations of topological objects are guided by the consideration of concrete manipulations that would be performed on rubber or other deformable material. Accordingly, experts have acquired a form of imagination that prompt them to re-draw diagrams and calculate with them, performing "epistemic actions" [22.20]. This form of imagination

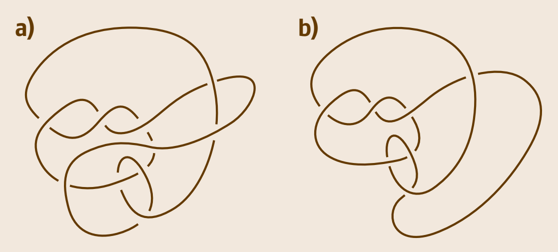

Fig. 22.10a,b Two nontrivial diagrams of the unknot

derives from our interaction with concrete objects and our familiarity with manipulating them. Moreover, the meaning of a knot diagram is fixed by its context of use: diagrams are the results of the interpretation of a figure, depending on the moves that are allowed on them and at the same time on the space in which they are embedded. Once the appropriate moves are established, the ambient space is fixed, thus determining the different equivalence relations. The context of use does not have to be predefined, preserving this kind of ambiguity that is not "damaging" [22.9], but productive. Actually, the indetermination of meaning makes different interpretations co-habit, and, therefore, allows attending to various properties and moves.

The same authors have also analyzed the practice of proving in low-dimensional topology [22.51]. As a case study, they have taken a specific proof: Rolfsen's demonstration of the equivalence of two presentations of the Poincaré homology sphere. This proof is taken from a popular graduate textbook: *Knots and Links* by *Rolfsen* [22.52]. The first presentation of Poincaré homology sphere is a *Dehn surgery*, while the second one is a *Heegaard diagram* (Fig. 22.11).

Without going into the details, the aim of the authors is to use this case study to show that, analogously to knot theory, *seeing* in low-dimensional topology means imagining a series of possible manipulations on the representations that are used, and is, of course, modulated by expertise. Moreover, the actual practice of proving in low-dimensional topology cannot be reduced to formal statements without loss of intuition. Several examples of representationally heterogeneous reasoning – that is neither entirely propositional nor entirely visual – are given. Both the very representations introduced and the manipulations allowed on them – what the authors, following a terminology proposed by *Larvor* [22.53], call *permissible actions* – are epistemologically relevant, since they are integral parts both of the reasoning and the justification provided. To claim that inferences involving visual representations are permissible only within a specific practice is to consider them as context dependent. A consequence would be that it is no longer

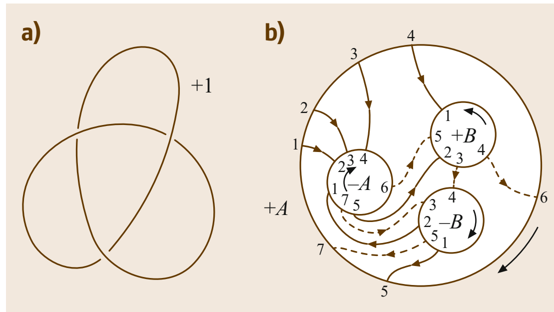

Fig. 22.11a,b The surgery code and the Heegaard diagram for the Poincaré homology sphere

possible to establish general criteria for mathematical validity, since they can only be local. The picture of mathematics emerging from these kinds of studies is thus very different from the one proposed from the post-nineteenth century philosophy of mathematics.

A final remark about representations in topology concerns a point about their materiality, already raised by Carter in a different context. To avoid confusion, it is necessary 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 see modifications on them, but experts may not find it necessary to actually draw all the physical pictures. The same holds for algebra where experts skip transitions that nontrained 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, which would determine what Thurston has called the *mental model*. To go back to Netz' analysis of the Euclidean diagram, here as well diagrams allow for procedures to be automatized or elided.

22.6 Computational Approaches

In this section, studies about the possibility of automatizing diagrammatic reasoning in mathematics are briefly introduced. Such attempts are worth being mentioned because they start from the observation that diagrammatic reasoning is crucial, at least in some areas of mathematics, and furthermore that any possible formalization for it should reflect its straightforwardness and directness. We will introduce the attempts of developing an automated reasoning program for plane geometry and for theory of numbers in turn.

22.6.1 (Manders') Euclid Reloaded

The analysis and the definitions provided by Manders about Euclidean geometrical reasoning were used to establish a formalization for diagrams in line with what he calls the diagram discipline. Such a project has brought about the creation of two logical systems, *E* [22.54] and *Eu* [22.55, 56], thanks to the work of *Avigad*, *Dean* and *Mumma*. Both systems produce formal derivations that line up closely with Euclid's proofs, in many cases following them step by step. (Another system that has been created to formalize Euclidean geometry is *FG* [22.57]. For details about *FG* and *Eu* and for a general discussion of the project in relation to model-based reasoning, see Chap. 23.). As summarized in a recent paper [22.58], the proof systems are designed to bring into sharp relief those attributes that are fundamental to Euclid's reasoning as characterized by Manders in his distinction between exact and co-exact properties. Nonetheless, the distinction is made with respect to a more restricted domain.

The Euclidean diagram has some components, which can be simple objects, such as points, lines, segments, and circles, and more complex ones, such as angles, triangles, and quadrilaterals. These components are organized according to some relations, which are the diagram attributes. Exact relations are obtained between objects having the same kind of magnitude: for example, for any two angles, the magnitude of one can be greater than the magnitude of the other or the same. Co-exact relations are instead positional: for example, a point can lie inside a region, outside it, or on its boundary. Co-exact relations concerning one-dimensional objects exclusively, such as line segments or circles, are intersection and nonintersection, while those concerning regions, one-dimensional or two-dimensional, are containment, intersection, and disjointness. Take the diagram in Fig. 22.12, representing the endpoint A as lying inside the circle H (a co-exact property), along with a certain distance between the point A and the circle's center B (an exact property). (Consider that in reproducing the diagram from Mumma's original article, the co-exact features were not affected, while the exact ones probably were.) Following Manders, in a proof in Euclid's system, premises and conclusions of diagrammatic inferences are composed of co-exact relations between geometric objects. In Fig. 22.13, an inference is shown (Fig. 22.13a) together with one of its possible associated diagrams (Fig. 22.13b).

In order to develop a formal system for these inferences, the main tasks in developing the programs were two: first, to specify the formal elements representing co-exact relations; and second, to formulate the rules in terms of the elements whereby diagrammatic inferences can be represented in derivations. The main difference between *Eu* and *E* is how the first task is modulated. *Eu* possesses a diagrammatic symbol type intended to model what is perceived in concrete physical diagrams, while *E* models the information directly extracted from concrete physical diagrams by providing a list of primitive relations recording co-exact information among three object types: points, lines, and circles. In Fig. 22.14, the formalization in *Eu* of the inference in Fig. 22.13a is shown. In Fig. 22.15, the formalization of the same inference in *E* is shown, with the primitive $on(\mathrm{A}, l)$ meaning *point* A *is on line* l.

In addition, the formalizations do not only have formal elements corresponding to Euclidean diagrams, but also formal elements corresponding to the Euclidean text, so as to also record exact information. In order to give a proof, the two kinds of representations have to interact.

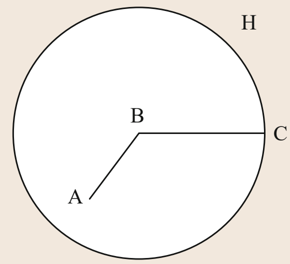

Fig. 22.12 A Euclidean diagram depicting exact and co-exact relations

a)

Points A and B are *on opposite sides* of line *l*
Points A and B are *on* line *m*

Line *m interects* line *l*

b)

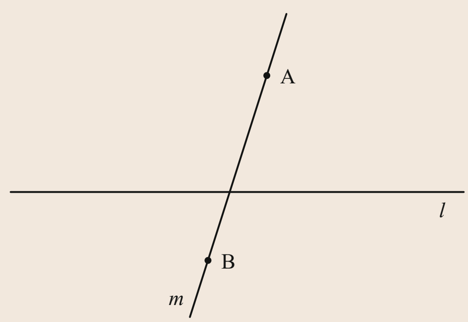

Fig. 22.13a,b An inference in Euclid's system according to Manders' reconstruction

22.6.2 Theorem Provers

Not only formalizations of Euclidean geometry have been provided. *Jamnik* developed a semi-automatic proof system, called DIAMOND (*Dia*gram*m*atic Reas*on*ing and *D*eduction), to formalize and mechanize diagrammatic reasoning in mathematics, and in particular to prove theorems of arithmetic using diagrams [22.59]. Interestingly, Jamnik starts by recording a simple cognitive fact, that is that given some basic mathematical training and our familiarity with spatial manipulations – remember the study on knot theory – it suffices to look at the diagram representing a theorem to understand not only what particular theorem it represents, but also that it constitutes a proof for it. As a consequence, one arrives at the belief that the theorem is correct. From here, the question is: Is it possible to simulate and formalize this kind of diagrammatic reasoning on machines? In other words, is this an example of intuitive reasoning that is particular to humans and machines are incapable of?

The first part of Jamnik's book provides a nice overview of the different diagrammatic reasoning systems that have been developed in the past century, such as, for example, *Gelernter*'s *Geometry Machine* [22.60] or *Koedinger* and *Anderson*'s *Diagram Configuration Model* [22.61]. For reasons of space, these systems will not be discussed here. In order to develop her proof system, she considers many different visual proofs in arithmetic and some of the analyses that have been given for them, by relying on the already mentioned collection edited by *Nelsen* [22.18, 19]. Such an analysis enables her to define a *schematic proof* as "a recursive function which outputs a proof of some proposition $P(n)$ given some n as input" [22.59, p. 52].

Consider inductive theorems with a parameter, which, in Jamnik's proposed taxonomy, are theorems where the diagram that is used to prove them represents one particular instance. An example of a theorem pertaining to this category is the *sum of squares of Fibonacci numbers*. According to this theorem, the sum of n squares of Fibonacci numbers equals the product of the n-th and $(n+1)$-th Fibonacci numbers. In symbols,

$$Fib(n+1) \times Fib(n) = Fib(1)^2 + Fib(2)^2 + \cdots + Fib(n)^2 \,. \tag{22.14}$$

The formal recursive definition of the Fibonacci numbers is given as

$$Fib(0) = 0\,, \quad Fib(1) = 1\,, \quad Fib(2) = 1\,,$$
$$Fib(n+2) = Fib(n+1) + Fib(n)\,. \tag{22.15}$$

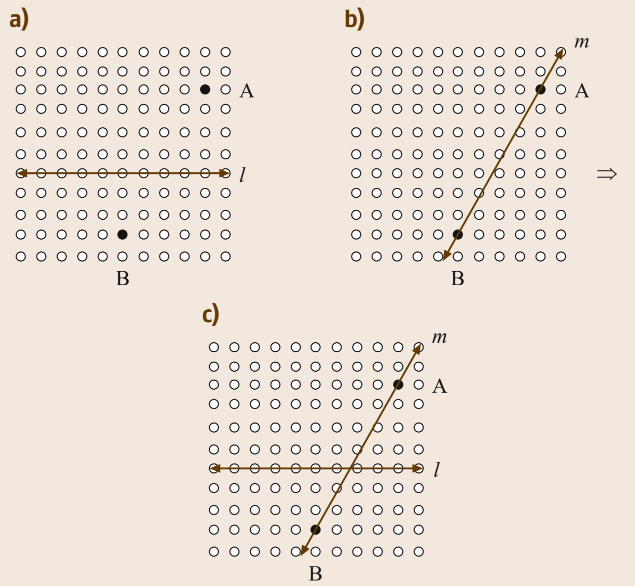

Fig. 22.14a–c The given inference in *EU*

a)	b)	c)
A, B points		
l line	on (A, m)	
Not *same side* (A, B, l)	on (B, m)	*intersects* (l, m)

Fig. 22.15a–c The given inference in *E*

Consider now Fig. 22.16. By looking at the spatial arrangement of the dots, we first take the rectangle of length $Fib(n+1)$ and height $Fib(n)$. Then, we split it in a square of magnitude $Fib(n)$, that is, the smaller side of the rectangle. We continue decomposing the remaining rectangle in a similar fashion until it is exhausted, that is, for all n. The sides of the created squares represent the consecutive Fibonacci numbers, and the longer side of every new rectangle is equal to the sum of the sides of two consecutive squares, which is precisely how the Fibonacci numbers are defined. As noted by Jamnik, the proof can also be carried out inversely, that is, starting from a square of unit magnitude $(Fib(1)^2)$ and joining it on one of its sides with another square of unit magni-

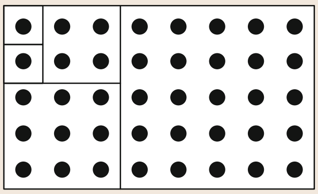

Fig. 22.16 Sum of squares of Fibonacci numbers

tude ($Fib(2)^2$): we have a rectangle. Then we can take the rectangle and join to it a square of the magnitude of its longer side, so as to create another rectangle. The procedure can be repeated for all n.

The schematic diagrammatic proof for this theorem would then be a sequence of steps that need to be performed on the diagram in Fig. 22.16:

1. Split a square from a rectangle. The square should be of a magnitude that is equal to the smaller side of a rectangle (note that aligning squares of Fibonacci numbers in this way is a method of generating Fibonacci numbers, that is, $1,1,1+1=2,1+2=3,2+3=5$, etc.).
2. Repeat this step on the remaining rectangle until it is exhausted.

These steps are sufficient to transform a rectangle of magnitude $Fib(n+1)$ by $Fib(n)$ to a representation of the right-hand side of the theorem, that is, n squares of magnitudes that are increasing Fibonacci numbers [22.59, p. 66].

A schematic proof is thus a schematic program which by instantiation at n gives a proof of every proposition $P(n)$. The constructive ω-rule justifies that such a recursive program is indeed a proof of a proposition for all n. This rule is based on the ω-rule, that is, an infinitary logical rule that requires an infinite number of premises to be proved in order to conclude a universal statement. The uniformity of this procedure is captured in the recursive program, for example, `proof`(n). Jamnik's attempt is thus to formalize and implement the idea that the generality of a proof is captured in a variable number of applications of geometrical operations on a diagram, and as a consequence to challenge the argument according to which human mathematical reasoning is fundamentally noncomputational, and, therefore, cannot be automatized. Details about DIAMOND's functioning cannot be given here. We just point out that also in this case diagrammatic reasoning is interpreted as a series of operations on a particular diagram, which can be repeated on other diagrams displaying the same geometric features.

22.7 Mathematical Thinking: Beyond Binary Classifications

The reader already acquainted with the topic of diagrammatic reasoning in mathematics might wonder why there has not yet been any explicit reference to the work of *Giaquinto*, who was undeniably one of the first philosophers to revive the attention toward mathematical visualization [22.62]. In one of his papers, we also find a nice overview of the literature concerning the possibility of obtaining rigorous proofs by reasoning diagrammatically [22.63]. The reason for this choice is that Giaquinto has not only been a pioneer in the renewed study of diagrammatic reasoning in mathematics, but also and even more interestingly he has given suggestions about the directions that future research should take. In this section, first his ideas on the role of visualization in mathematical discovery will be briefly presented, and then his proposal about how to consider mathematical thinking in general will be discussed (in the course of the final revisions of the present chapter, I discovered that *Giaquinto* recently published an entry on a topic that is related to ours, see for reference [22.64]).

As Giaquinto makes it clear, his original motivation for studying visual thinking in mathematics was to provide an epistemology of individual discovery and of actual mathematical thinking, so as to reopen the investigation of early thinkers from Plato to Kant, who indeed had as an objective to explore the nature of the individual's basic mathematical beliefs and skills. His strategy is thus first to acknowledge that there is more than one kind of thinking in mathematics, and then to assess the epistemic status of each of these kinds of mathematical thinking. For this reason, and as he himself admits, his view is in some sense much more *traditional* than many of the works produced by the post-nineteenth philosophers of mathematics. Discovery is a very crucial issue for the practice of mathematics and another topic that unfortunately has been neglected by post-nineteenth century approaches, which focused mainly on logic, proof, and justification. Giaquinto tries to give an account of the complexity of mathematical thinking, and to this aim he also inquires into fields of research other than philosophy, thus trespassing disciplinary boundaries. His belief is that cognitive science constitutes a new tool that can be helpful for understanding mathematical thinking: though cognition has always been the object of philosophy, the development of cognitive science surely represents an advantage for the philosophers of our century over the scholars of earlier times. Another discipline that could be an ally in disclosing mathematical thinking is mathematical education, traditionally categorized as an applied field unable to provide conclusive hints for theoretical research. Moreover, Giaquinto assigns an important role to history, both the history of mathematics and the history of philosophy.

The main epistemological thesis of the book is that there is no reason to assume a uniform evaluation that would fit all cases of visual thinking in mathematics, since visual operations are diverse depending on the mathematical context. Moreover, in order to assess this thesis, we do not need to refer to advanced mathematics: basic mathematics is already enough to account for the process of mathematical discovery by an individual who reasons visually. In fact, only the final part of the book goes beyond very elementary mathematics.

It should be mentioned that also Giaquinto defends a neo-Kantian view according to which in geometry we can find cases of synthetic a priori knowledge, that is cases that do not involve either analysis of meanings or deduction from definitions. In fact, he refers to the already mentioned study by *Norman*, which is neo-Kantian in spirit, as a strong case showing that following Euclid's proof of the proposition that the internal angles of a triangle sum to two right angles require visual thinking, and that visual thinking is not replaceable by nonvisual thinking [22.23]. Nonetheless, the focus in this section will be mostly on the last chapter of the book, where Giaquinto discusses how the traditional twofold division between *algebraic* thinking versus *geometric* thinking is not appropriate for accounting for mathematical reasoning. His conclusion, which can be borrowed also as a conclusion of the present chapter, is that there is a need for a much more comprehensive taxonomy for spatial reasoning in mathematics, which that would include operations such as visualizing motion, noticing reflection symmetry, and shifting aspects. In fact, if one considers thinking in mathematics as a whole, then there arises a sense of dissatisfaction with any of the common binary distinctions that have been proposed between algebraic thinking on the one hand and geometric thinking on the other; the philosopher's aim should be to move toward a much more discriminating taxonomy of kinds of mathematical reasoning.

Consider, for example, *aspect shifting* as precisely one form of mathematical thinking that seems to elude standard distinctions. Aspect shifting is the same cognitive ability that Macbeth describes in discussing the way in which the Greek geometer – and every one of us today who practices Euclidean geometry – reasoned *in* the Euclidean diagram. Take again the visual proof given in Sect. 22.1 for the Pythagorean theorem. As *Giaquinto* explains, it is possible to look at the square in Fig. 22.1b – that has letters in it, and, therefore, is a kind of *lettered* diagram – and *see* that the area of the larger square is equal to the area of the smaller square plus the area of the four right-angled triangles [22.62, pp. 240–241]. How do we acquire this belief? Giaquinto's reply is that first we have to reason *geometrically* and shift between aspects, so as to recognize that the area of the square is both $(a+b)^2$ and $2ab+c^2$. From here, we then have to proceed *algebraically* as follows

$$a^2+2ab+b^2=2ab+c^2\,,\quad a^2+b^2=c^2\,. \qquad (22.16)$$

At this point, by looking back at the figure, we realize – *geometrically* again – that the smaller square is also the square of the hypotenuse of the right-angled triangle. Finally, from the formula, we conclude that the area of the square of the hypotenuse is equal to the sum of the squares of its other two sides. Then the question is: Is this argument as a whole to be considered as primary algebraic or geometric? It seems that neither of these two categories would be fully appropriate to capture it.

This is an interesting point also relative to other kinds of mathematical reasoning by means of some particular representation. Consider a notation that is used in topology and take as an example the *torus* that can be defined as a square with its sides identified. In order to obtain the torus from a square, we identify all its four sides in pairs. The square in Fig. 22.17a has arrows in it indicating the gluings, that is, the identifications. First, we identify two sides in the same direction, so as to obtain the *cylinder* (Fig. 22.17b); then, we identify the other two, again in the same direction: in Fig. 22.17c, one can see the torus with two marked curves, where the gluings, that is, the identifications, were made.

In discussing the role of notation in mathematics, *Colyvan* takes into consideration diagrams such as the one in Fig. 22.17a, and points out that this notation is "something of a halfway house between pure algebra and pure geometry" [22.65, p. 163]. In Colyvan's view these diagrams are, on the one hand, a piece of notation, but, on the other, also an indication on how to construct the object in question. The first feature seems to belong to algebra, while the second to geometry. Moreover, note that if we identify two sides of the square

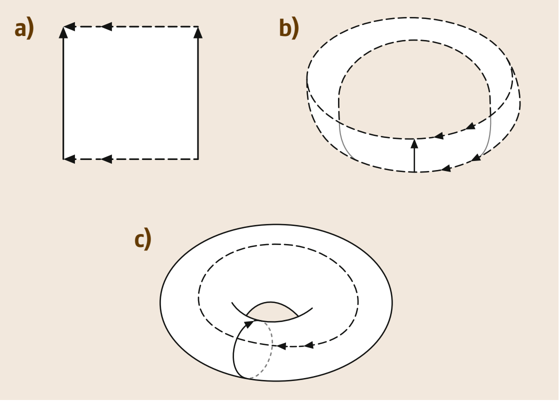

Fig. 22.17a–c Constructing the torus

in the same direction and the other two in the opposite direction, we obtain the *Klein bottle*, which is a very peculiar object, since it is three-dimensional but needs four spatial dimensions for its construction, and, even more interestingly, has no inside or outside. The Klein bottle demonstrates how powerful such a notation is: it leads to objects that would be otherwise considered as nonsense, and it also allows us to deduce their properties. As *Colyvan* summarizes, "Whichever way you look at it, we have a powerful piece of notation here that does *some genuine mathematical work for us*" ([22.65, p. 163], emphasis added). Diagrams, as well as other *powerful* notations, operate *at our place*. Moreover, at least some of them seem to be some kinds of *hybrid* objects, trespassing boundaries. They are geometric and algebraic at the same time.

Consider again the relations between the algebra of combinatorial groups and their geometry (Sect. 22.5.2). As Starikova tells us, first the combinatorial group theory was amplified with a geometric element – a graph – where *geometric* refers mostly to geometric constructions as methods of geometry rather than algebra. But eventually this geometric element was significantly expanded and groups became geometric objects in virtue of their revealed geometric properties. The introduction of graphs thus provided mathematicians with a powerful instrument for facilitating their intuitive capacities and furthermore gave a good start for further intuitions which finally lead to advanced conceptual links with geometry and the definition of a broader geometric arsenal to algebra. Also in the knot theory example (Sect. 22.5.3), knot diagrams are shown to have at the same time diagrammatic and symbolic elements, and, as a consequence, their nature cannot be captured by the traditional dichotomy between geometric and algebraic reasoning. All this is to show that *Giaquinto's* invitation to define a "more discriminating and more comprehensive" [22.13, p. 260] taxonomy for mathematical thinking going beyond twofold divisions is still valid, and that more on this topic needs to be done.

22.8 Conclusions

The objective of the present chapter was to introduce the different studies that have recently been devoted to diagrammatic reasoning in mathematics. The first topic discussed was the role of diagrams in Euclidean and Greek geometry in general (Sect. 22.3); then, the productive ambiguity of diagrams was defined (Sect. 22.4) and case studies in contemporary mathematics were briefly reviewed (Sect. 22.5). It has been shown how some attempts have tried to automatize diagrammatic reasoning in mathematics, in particular to formalize arguments in Euclidean geometry and proofs in theory of numbers (Sect. 22.6); finally, it has been argued that the attention to diagrammatic reasoning in mathematics can shed light on the fact that mathematics makes use of different kinds of representations that are so intertwined that it is difficult to draw sharp distinctions between the different subpractices and the corresponding reasoning (Sect. 22.7). We started from the study of diagrammatic reasoning and we arrived at the consideration of mathematical thinking as a whole, and of the role of notations and representations in it. Mathematicians use a vast range of cognitive tools to reason and communicate mathematical information; some of these tools are material, and, therefore, they can easily be shared, inspected, and reproduced. Specific representations are introduced in a specific practice and, once they enter into the set of the available tools, they may have an influence on the very same practice. This process plays a significant role in mathematics.

There is a last remark to make at the end of this survey, that is, that in diagrammatic reasoning, we have seen the continuity and the discreteness of space operating. Continuous was the space of the Euclidean diagram, discrete (at least in part) the space of the diagrams for Galileo's theorem and for the sum of the Fibonacci numbers. Diagrammatic reasoning thus seems to have fundamentally a geometric nature, since it organizes space. Nonetheless, we have also shown that a diagram never comes alone, but always with some form of text giving indications for its construction or stipulating its correct interpretation. The relation between the diagram and text is defined each time by the specific practice. As a consequence, diagrams appear to be very interesting hybrid objects, whose nature cannot be totally captured by standard oppositions. They are cognitive tools available for thought, whose effectiveness depends on both our spatial and our linguistic cognitive nature.

Acknowledgments. My thanks go to the people quoted in the text, for their work and for the fruitful discussions in which I have taken part at recent conferences and workshops. I am particularly indebted to Mario Piazza and Silvia De Toffoli, with whom I have extensively reflected upon the topic of diagrammatic reasoning in mathematics. I am also grateful to Albrecht Heeffer, who gave me the occasion of working on this chapter and to the *Université de Lorraine* and the *Région Lorraine* for having sustained my research.

References

22.1 J.R. Brown: *Philosophy of Mathematics: An Introduction to the World of Proofs and Pictures* (Routledge, New York 1999)
22.2 D. Sherry: The role of diagrams in mathematical arguments, Found. Sci. **14**, 59–74 (2009)
22.3 S.-J. Shin: Heterogeneous reasoning and its logic, Bull. Symb. Log. **10**(1), 86–106 (2004)
22.4 E. Maor: *The Pythagorean Theorem. A 4000-Year History* (Princeton Univ. Press, Princeton 2007)
22.5 J. Høyrup: Tertium non datur: On reasoning styles in early mathematics. In: *Visualization, Explanation and Reasoning Styles in Mathematics*, *Synthese Library*, Vol. 327, ed. by P. Mancousu, K.F. Jørgensen, S.A. Pedersen (Springer, Dordrecht 2005) pp. 91–121
22.6 K. Chemla: The interplay between proof and algorithm in 3rd century China: The operation as prescription of computation and the operation as argument. In: *Visualization, Explanation and Reasoning Styles in Mathematics*, ed. by P. Mancosu, K.F. Jørgensen, S.A. Pedersen (Springer, Berlin 2005) pp. 123–145
22.7 K. Stenning, O. Lemon: Aligning logical and psychological perspectives on diagrammatic reasoning, Artif. Intell. Rev. **15**, 29–62 (2001)
22.8 J. Barwise, J. Etchemendy: Visual information and valid reasoning. In: *Logical Reasoning with Diagrams*, ed. by G. Allwein, J. Barwise (Oxford Univ. Press, Oxford 1996) pp. 3–25
22.9 S.-J. Shin, O. Lemon, J. Mumma: Diagrams. In: *The Stanford Encyclopedia of Philosophy*, ed. by E. Zalta, Fall 2013 Edition, http://plato.stanford.edu/archives/fall2013/entries/diagrams/
22.10 S.-J. Shin: The mystery of deduction and diagrammatic aspects of representation, Rev. Philos. Psychol. **6**, 49–67 (2015)
22.11 B. Russell: *The Principles of Mathematics* (W.W. Norton, London 1903/ 1937)
22.12 R. Netz: *The Shaping of Deduction in Greek Mathematics: A Study of Cognitive History* (Cambridge Univ. Press, Cambridge 1999)
22.13 M. Giaquinto: *The Search for Certainty* (Oxford Univ. Press, Oxford 2002)
22.14 F. Klein: *Elementary Mathematics from an Advanced Standpoint* (Dover, Mineola 2004), the first German edition is 1908
22.15 D. Hilbert: *The Foundations of Geometry* (K. Paul, Trench, Trübner, London 1899/ 1902)
22.16 P. Mancosu, K.F. Jørgensen, S.A. Pedersen (Eds.): *Visualization, Explanation and Reasoning Styles in Mathematics* (Springer, Berlin 2005)
22.17 V.F.R. Jones: A credo of sorts. In: *Truth in Mathematics*, ed. by H.G. Dales, G. Oliveri (Clarendon, Oxford 1998)
22.18 R. Nelsen: *Proofs without Words II: More Exercises in Visual Thinking*, Classroom Resource Materials (The Mathematical Association of America, Washington 2001)
22.19 R. Nelsen: *Proofs without Words: Exercises in Visual Thinking*, Classroom Resource Materials (The Mathematical Association of America, Washington 1997)
22.20 D. Kirsh, P. Maglio: On distinguishing epistemic from pragmatic action, Cogn. Sci. **18**, 513–549 (1994)
22.21 J. Ferreiros: *Mathematical Knowledge and the Interplay of Practices* (Princeton Univ. Press, Princeton 2015)
22.22 L.A. Shabel: *Mathematics in Kant's Critical Philosophy: Reflections on Mathematical Practice* (Routledge, New York 2003)
22.23 J. Norman: *After Euclid* (CSLI Publications, Univ. Chicago Press, Chicago 2006)
22.24 C.S. Peirce: *Collected Papers* (The Belknap Press of Harvard Univ. Press, Cambridge 1965)
22.25 J. Azzouni: Proof and ontology in Euclidean mathematics. In: *New Trends in the History and Philosophy of Mathematics*, ed. by T.H. Kjeldsen, S.A. Pederson, L.M. Sonne-Hansen (Univ. Press of Southern Denmark, Odense, Denmark 2004) pp. 117–133
22.26 W.P. Thurston: On proof and progress in mathematics, Bull. Am. Math. Soc. **30**(2), 161–177 (1994)
22.27 P. Mancosu (Ed.): *The Philosophy of Mathematical Practice* (Oxford Univ. Press, Oxford 2008)
22.28 K. Manders: The Euclidean diagram. In: *The Philosophy of Mathematical Practice*, ed. by P. Mancosu (Oxford Univ. Press, Oxford 2008) pp. 80–133
22.29 K. Manders: Diagram-based geometric practice. In: *The Philosophy of Mathematical Practice*, ed. by P. Mancosu (Oxford Univ. Press, Oxford 2008) pp. 65–79
22.30 D. Macbeth: Diagrammatic reasoning in Euclid's elements. In: *Philosophical Perspectives on Mathematical Practice*, Vol. 12, ed. by B. Van Kerkhove, J. De Vuyst, J.P. Van Bendegem (College Publications, London 2010)
22.31 H.P. Grice: Meaning, Philos. Rev. **66**, 377–388 (1957)
22.32 P. Catton, C. Montelle: To diagram, to demonstrate: To do, to see, and to judge in Greek geometry, Philos. Math. **20**(1), 25–57 (2012)
22.33 D. Macbeth: Diagrammatic reasoning in Frege's *Begriffsschrift*, Synthese **186**, 289–314 (2012)
22.34 M. Panza: The twofold role of diagrams in Euclids plane geometry, Synthese **186**(1), 55–102 (2012)
22.35 M. Panza: Rethinking geometrical exactness, Hist. Math. **38**, 42–95 (2011)
22.36 C. Parsons: *Mathematical Thought and Its Objects* (Cambridge Univ. Press, Cambridge 2008)
22.37 P. Mancosu (Ed.): *From Brouwer to Hilbert. The Debate on the Foundations of Mathematics in the 1920s* (Oxford Univ. Press, Oxford 1998)
22.38 Proclus: *In primum Euclidis Elementorum librum commentarii* (B.G. Teubner, Leipzig 1873), ex recognitione G. Friedlein, in Latin
22.39 Proclus: *A Commentary on the First Book of Euclid's Elements* (Princeton Univ. Press, Princeton 1992), Translated with introduction and notes by G.R. Morrow
22.40 Aristotle: *Metaphysics*, Book E, 1026a, 6–10
22.41 E. Grosholz: *Representation and Productive Ambiguity in Mathematics and the Sciences* (Oxford

Univ. Press, Oxford 2007)
22.42 K. Chemla: Lazare Carnot et la Généralité en Géométrie. Variations sure le Théoréme dit de Menelaus, Rev. Hist. Math. **4**, 163–190 (1998), in French
22.43 J. Carter: Diagrams and proofs in analysis, Int. Stud. Philos. Sci. **24**(1), 1–14 (2010)
22.44 J. Carter: The role of representations in mathematical reasoning, Philos. Sci. **16**(1), 55–70 (2012)
22.45 U. Haagerup, S. Thorbjørnsen: Random matrices and K-theory for exact C^*-algebras, Doc. Math. **4**, 341–450 (1999)
22.46 M.E. Moore: *New Essays on Peirce's Mathematical Philosophy* (Open Court, Chicago and La Salle 2010)
22.47 K. Manders: Euclid or Descartes: Representation and responsiveness, (1999), unpublished
22.48 I. Starikova: Why do mathematicians need different ways of presenting mathematical objects? The case of Cayley graphs, Topoi **29**, 41–51 (2010)
22.49 I. Starikova: From practice to new concepts: Geometric properties of groups, Philos. Sci. **16**(1), 129–151 (2012)
22.50 S. De Toffoli, V. Giardino: Forms and roles of diagrams in knot theory, Erkenntnis **79**(4), 829–842 (2014)
22.51 S. De Toffoli, V. Giardino: An inquiry into the practice of proving in low-dimensional topology. In: *From Logic to Practia*, (Springer, Cham 2015) pp. 315–336
22.52 D. Rolfsen: *Knots and Links* (Publish or Perish, Berkeley 1976)
22.53 B. Larvor: How to think about informal proofs, Synthese **187**(2), 715–730 (2012)
22.54 J. Avigad, E. Dean, J. Mumma: A formal system for Euclid's elements, Rev. Symb. Log. **2**(4), 700–768 (2009)
22.55 J. Mumma: Proofs, pictures, and Euclid, Synthese **175**(2), 255–287 (2010)
22.56 J. Mumma: Intuition formalized: Ancient and modern methods of proof in elementary geometry, Ph.D. Thesis (Carnegie Mellon University, Pittsburgh 2006)
22.57 N. Miller: *Euclid and His Twentieth Century Rivals: Diagrams in the Logic of Euclidean Geometry* (CSLI Publications, Stanford 2007)
22.58 Y. Hamami, J. Mumma: Prolegomena to a cognitive investigation of Euclidean diagrammatic reasoning, J. Log. Lang. Inf. **22**, 421–448 (2013)
22.59 M. Jamnik: *Mathematical Reasoning with Diagrams* (Univ. Chicago Press, Chicago 2002)
22.60 H. Gelernter: Realization of a geometry theorem-proving machine. In: *Computers and Thought*, ed. by E. Feigenbaum, J. Feldman (Mac Graw Hill, New York 1963) pp. 134–152
22.61 K.R. Koedinger, J.R. Anderson: Abstract planning and perceptual chunks, Cogn. Sci. **14**, 511–550 (1990)
22.62 M. Giaquinto: *Visual Thinking in Mathematics* (Oxford Univ. Press, Oxford 2007)
22.63 M. Giaquinto: Visualizing in mathematics. In: *The Philosophy of Mathematical Practice*, ed. by P. Mancosu (Oxford Univ. Press, Oxford 2008) pp. 22–42
22.64 M. Giaquinto: The epistemology of visual thinking in mathematics. In: *The Stanford Encyclopedia of Philosophy*, ed. by E.N. Zalta, Winter 2015 Edition, http://plato.stanford.edu/archives/win2015/entries/epistemology-visual-thinking/
22.65 M. Colyvan: *An Introduction to the Philosophy of Mathematics* (Cambridge Univ. Press, Cambridge 2012)