2

What You See Is What You Get

2.1 The starting point: mirrors and reflections

I use, as one of the principal running themes of this book, a comparison between two approaches to the concept of symmetry as it is understood and used in modern algebra. The corresponding mathematical discipline is well established and is called the *theory of finite reflection groups*. The reader should not worry if he or she has never encountered this name; as will soon be seen, the subject has many elementary facets.

To start with, the principal objects of the theory can be defined in the most intuitive way. First I give an informal description:

> Imagine a few (semi-transparent) mirrors in ordinary three-dimensional space. Mirrors (more precisely, their images) multiply by reflecting in each other, as in a kaleidoscope or a gallery of mirrors. Of special interest are mirror systems of which generate only finitely many reflected images. Such finite systems of mirrors happen to be one of the cornerstones of modern mathematics and lie at the heart of many mathematical theories.

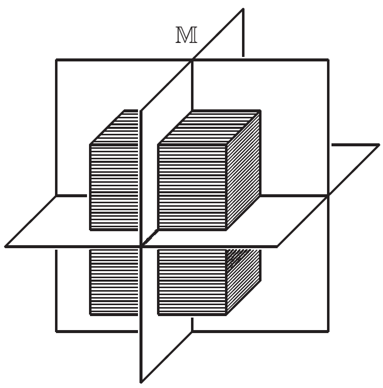
As usual, the full theory is concerned with the more general case of n-dimensional Euclidean space, with 2-dimensional mirrors replaced by $(n-1)$-dimensional *hyperplanes*. To that end, we give a formal definition:

> A system of hyperplanes (mirrors or images of mirrors) $\mathbb{M}$ in the Euclidean space $\mathbb{R}^n$ is called *closed* if, for any two mirrors M_1 and M_2 in $\mathbb{M}$, the mirror image of the mirror M_2 in the mirror M_1 also belongs to $\mathbb{M}$ (Figure 2.1).

Anna Borovik,
née Vvedenskaya,
aged 8

Thus, the principal objects of the theory are *finite closed systems of mirrors*. In more evocative terms, the theory can be described as the geometry of *multiple mirror images*. This approach to symmetry is well known and is found, for example, in Chapter 5, §3 of Bourbaki's classical text [323][1], or in Vinberg's paper [415]; I have recently used it in my textbook *Mirrors and Reflections* [294].

However, closed systems of mirrors are usually known in mathematics under a different name, and in a completely different dress, as *finite reflection groups*. They make up a classical chapter of mathematics, which originated in the seminal works of H. S. M. Coxeter [336, 337] (hence yet another name: *finite Coxeter groups*). The theory can be based on the concept of a group of transformations (as is done in many excellent books; see, for example, [360, 367]) and can be developed in group-theoretic terms.



The system $\mathbb{M}$ of all mirrors of symmetry of a geometric body Δ is *closed*: the reflection of a mirror in another mirror is yet another mirror. Notice that if Δ is compact (i.e., closed and bounded), then all mirrors have a point in common.

Fig. 2.1. A closed system of mirrors. Drawing by Anna Borovik.

So we have two treatments, in two different mathematical languages, of the same mathematical theory (which I will call *Coxeter theory*). This is by no means an unusual thing in mathematics. What makes mirror systems/Coxeter groups interesting is that a closer look at the corresponding mathematical languages reveals their cognitive (and even neurophysiological!) aspects, much more obviously than in the rest of mathematics. In particular, as we shall soon see, the mirror system/Coxeter group alternative precisely matches the great *visual* / *verbal* divide of mathematical cognition.

It is worthwhile to pause for a second over the question of why we pay special attention to visual and speech processing. The answer is obvious: of all our senses, sight and hearing have the highest information processing rate and are used for communication. One can only speculate what mathematics would look like if we had an echolocating capacity (see a discussion of the way bats perceive the world in a paper by Kathleen Akins [146]). This is even more mind-boggling: try to imagine that humans have electric sensing and communicating facilities of the kind that Nile elephant fish *Gymnarchus niloticus* have and therefore that we live in a landscape made not of shapes and volumes (sight is of no use in the murky waters of the Nile) but of electromagnetic capacities and conductivities.[2] Would the concepts and results of vector calculus be self-evident to us? And a further question can be asked: which immediately intuitive mathematical concepts would become less intuitive? In Section 4.4 I attempt to suggest a partial answer to this question.

I wish to stress that, although the theory of Coxeter groups formally belongs to "higher" mathematics, the issues raised in the next two chapters are relevant to the teaching and understanding of mathematics at all levels, from elementary school to graduate studies. Indeed, I will be talking about such matters as *geometric intuition*. I will also touch on the role of pictorial proofs and self-explanatory diagrams; some of these may seem naïve, but, as I hope to demonstrate, they frequently lead deep into the heart of mathematics (see Section 2.6 for one of the more striking cases).

2.2 Image processing in humans

The mirror is one of the most powerful and evocative symbols of our culture; seeing oneself in a mirror is equated to self-awareness. But the reason why the language of mirrors and reflections happens to be so useful in the exposition of mathematical theories lies not so much at a cultural as at a psychophysiological level.

Erich Ellers, aged 7

How do people recognize mirror images? Tarr and Pinker [235] showed that recognition of mirror images of planar shapes is done by subconscious mental rotation of 180° about an appropriately chosen axis. Remarkably, the brain computes the position of this axis!

This is how Pinker describes the effect of their simple experiment.

> So we showed ourselves [on a computer screen] the standard upright shape alternating with one of its mirror images, back and forth once a second. The perception of flipping was so obvious that we didn't bother to recruit volunteers to confirm it. When the shape alternated with its upright reflection, it seemed to pivot like

> a washing machine agitator. When it alternated with its upside-down reflection, it did backflips. When it alternated with its sideways reflection, it swooped back and forth around the diagonal axis, and so on. *The brain finds the axis every time*. [218, pp. 282–283]

Interestingly, the brain does exactly the same with randomly positioned three-dimensional shapes, provided they *have the same chirality* (that is, both are of left-hand or right-hand type, as gloves, say) and can be identified by a rotation [233]. The interested reader may wish to take any computer graphics package which allows animation and see it for himself or herself.[3]

In view of these experiments, it is difficult to avoid the conclusion that Euler's classical theorem is hardwired into our brains:

> If an orientation-preserving isometry of the affine Euclidean space $\mathbb{AR}^3$ has a fixed point, then it is a rotation around some axis.

This illusion of rotation disappears when the brain faces the problem of the identification of three-dimensional mirror images of *opposite chirality*; indeed, they can still be identified by an appropriate rotation, but, this time, in four-dimensional space. The environment which directed the evolution of our brain never provided our ancestors with four-dimensional experiences.

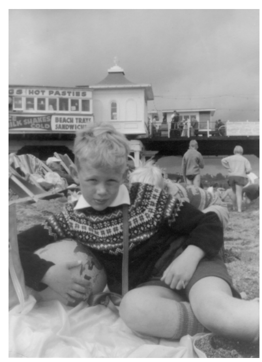
It is difficult to avoid the conclusion that Euler's Theorem is hardwired into our brains.

Human vision is a solution of an ill-posed inverse problem of recovering information about three-dimensional objects from two-dimensional projections on the retinas of the eyes. Pinker stresses that this problem is solvable only because of the multitude of assumptions about the nature of the objects and the world in general built into the human brain or acquired from previous experiences.[4]

The algorithm of the identification of three-dimensional shapes is only one of many modules in the immensely complex system of visual processing in humans. It is likely that various modules are implemented as particular patterns of connections between neurons. It is natural to assume that different modules developed at different stages of evolution [234]. The older ones are likely to be simpler and involve relatively simple wiring diagrams. But since they had adaptive value, they were inherited and they acted as constraints in the evolution of later additions to the system, in particular any new modules which happened to process the outputs of, and interact with, the pre-existent modules. At every stage, evolution led to the development of an algorithm for solving a very special and narrow problem. Of course, the evolution is guided by the universal and basic principle of the survival of the fittest. But, translated into selection criteria for the gradual improvement of light-sensing organs, the general principle became highly specialized and changing over time. First it favored higher sensitivity to diffuse light of a few cells which previously had quite different functions; at the next step of evolution it favored individuals with light-sensitive cells positioned in a more efficient way and, most probably, only after that, started to favor individuals who had some mechanism for discriminating between light stimuli applied to different groups of cells. Therefore we should expect that the image processing algorithms of our brain have a multilayered structure which reflects their ontogenesis—and, not unlike many modern software systems, are full of outdated "legacy code".

The "flipping" algorithm for the recognition of mirror images of a flat object and the closely related (and possibly identical) "rotation" algorithm for making randomly orientated three-dimensional objects coincide provide rare cases where we can glimpse the inner workings of our mind. Observe, however, that the algorithms are solutions of relatively simple mathematical problems with a very rigid underlying mathematical structure, namely, the group of isometries of three-dimensional Euclidean space. There is no analogue of Euler's Theorem for four-dimensional space![5]

The reader has possibly noticed that I prefer to use the term "algorithm" rather than "circuit", emphasizing the strong possibility that a given algorithm can be implemented by different circuit arrangements if some of the arrangements become impossible as the result of trauma, especially during the early stages of a child's development.



David Broomhead, aged 8

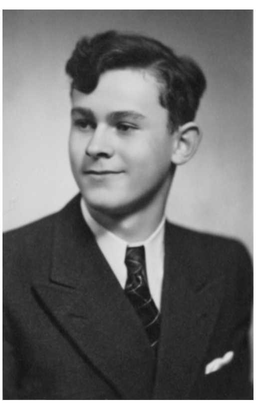
Studies of compensatory developments are abundant in the literature. When I was looking for some recent studies, my colleague David Broomhead directed me to the paper [182], a case study of a young woman who has been unable to make eye movements since birth but has surprisingly normal visual perception. This is astonishing because the so-called saccadic movements of the eyes are crucial for tracing the contours and the key features of objects. Try to experiment with a mirror: you will not see your eyes moving. During each saccade, the eye is in effect blind. We see the world frame-by-frame, as in the cinema. The continuity of the moving world is the result of the work of sophisticated interpolating routines integrated into the visual processing modules of our brain. Not surprisingly, continuity is one of the most intuitive (although hard to formalize) concepts of mathematics.

The woman in the study reported in [182] compensates for her lack of eye movement by quick movements of her head which follow the usual highly regular patterns of saccadic movements. I quote from the paper: “Her case suggests that saccadic movements, of the head or the eye, form the *optimal sampling method* for the brain”. The italics are mine, since I find the choice of words very suggestive: mathematics is encroaching on the inner working of the brain, raising some really interesting metamathematical questions.

2.3 A small triumph of visualization: Coxeter’s proof of Euler’s Theorem

If you need convincing that visualization is a purposeful tool in learning, teaching, and doing mathematics, there is no better example than the proof of Euler’s Theorem as given by Coxeter [338, p. 36]; I quote it *verbatim*. Remember that Coxeter’s book was first published in 1948, so it was written for readers who were likely to have taken a standard course of Euclidean geometry and therefore had developed their geometric imagination.

> In three dimensions, a congruent transformation that leaves a point **O** invariant is the product of at most three reflections: one to bring together the two x-axes, another for the y-axes, and a third (if necessary) for the z-axes.
>
> Since the product of three reflections is opposite, a direct transformation with an invariant point **O** can only be the product of reflections in *two* planes through **O**, i.e., a rotation.



Erich Ellers, aged 15

I add just a few comments to facilitate the translation into modern mathematical language: a *congruent transformation* is an isometry; a *direct transformation* preserves the orientation (chirality), while an *opposite* transformation changes it. Coxeter refers to the fact that the product of two mirror reflections is a rotation about the line of intersection of the mirrors. This is something that everyone has seen in a tri-fold dressing table mirror; the easiest way to prove the fact is to notice that the product of two reflections leaves invariant every point on the line of intersection of the mirrors.[6]

We humans are blessed with a remarkable piece of mathematical software for image processing directly hardwired into our brains. Coxeter made full use of it and expected the reader to use it, in his lightning proof of Euler’s Theorem. (See a further discussion of Coxeter’s proof in Section 6.3.) The perverse state of modern mathematics teaching is that “geometric intuition”, the skill of solving geometric problems by looking at (simplified) two- and three-dimensional models, has been largely expelled from the classroom practice.[7]

However, our geometric intuition involves at least two quite different (although closely related) cognitive components: visual processing and motor control. The latter is paradoxical; our hands can move and act with extreme precision, but we receive much less information feedback from the feeling of the motion itself or from the position of our body and hands.

The perverse state of modern mathematics teaching is that "geometric intuition" has been largely expelled from classroom practice.

To illustrate mathematical implications of this difference, I offer a small problem directly related to Euler's Theorem. I quote it from the book by David Henderson and Daina Taimina [272], where it is discussed in a slightly different context:

> When grinding a precision flat mirror, the following method is sometimes used: Take three approximately flat pieces of glass and put pumice between the first and second pieces and grind them together. Then do the same for the second and the third pieces and then for the third and first pieces. Repeat many times and all three pieces of glass will become very accurately flat.

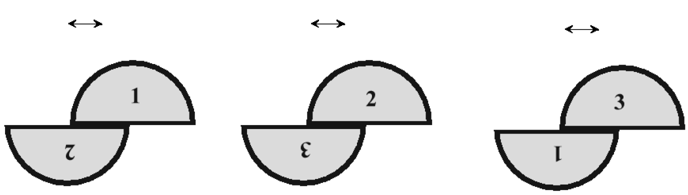


Fig. 2.2. Grinding a plane mirror (after David Henderson and Daina Taimina [272]).

See Figure 2.2. Now close your eyes and try to imagine your hands gently sliding one piece of glass all over the other. Do you see why this works?

Now I separate the question into two sub-questions, which, I believe, refer to two different levels of our intuition.

Indeed, why?

(A) Why do we need *three* pieces of glass to achieve perfect flatness? [?]

Answer it!

(B) Here is a trickier question: if only two pieces of glass are used, and the resulting surface is not plane, then (assuming that the grinding was thorough and even) what is this surface? [?]

It is futile to talk about mathematical practice without first acknowledging that it can only be understood alongside its interaction with the human brain.

The reader may wish to ponder these questions for a while; I give the answers in Section 4.5.

I am writing this book because I believe it is futile to talk about mathematical practice without first acknowledging that mathematics is an activity of the human mind and, in particular, the human brain. But our mind—or our cognitive system—is not homogeneous: its different parts developed at different stages of evolution, they have different levels of sophistication, an interaction between different modules is frequently awkward. We will not get much understanding of how mathematics lives in our minds without taking into account all the complexities and limitations of its constituent parts.

2.4 Mathematics: interiorization and reproduction

What is Mathematics, Really?
Reuben Hersh [48]

But Didactylos posed the famous philosophical conundrum:

*"Yes, But What's It **Really** All About, Then,*
When You Get Right Down To It,
*I **Mean** Really?"*

Terry Pratchett [434, p. 167]

I have already quoted Davis and Hersh [21, p. 399], to say that mathematics is

> *the study of mental objects with reproducible properties.*

A famous mathematician, David Mumford, uses this formulation in his paper [72, p. 199] and further comments on it:

> I love this definition because it doesn't try to limit mathematics to what has been called mathematics in the past but really attempts to say why certain communications are

> classified as math, others as science, others as art, others as gossip. Thus reproducible properties of the physical world are science whereas reproducible mental objects are math. Art lives on the mental plane (the real painting is not the set of dry pigments on the canvas nor is a symphony the sequence of sound waves that convey it to our ear) but, as the post-modernists insist, is reinterpreted in new contexts by each appreciator. As for gossip, which includes the vast majority of our thoughts, its essence is its relation to a unique local part of time and space.

If we accept this definition of mathematics, then we have to address two intertwined aspects of learning and mastering mathematics:

- the development of reproduction techniques for our own mental objects,
- interiorization of other people's mental objects.

There is a natural hierarchy of the methods of reproduction. A partial list in roughly descending order includes: proof; axiomatization; algorithm; symbolic and graphic expression. I wish to make it clear that reproduction is more than communication: you have to be able to reproduce your own mental work *for yourself*. Maybe it even makes sense to view *recovery procedures* for lost or forgotten mathematical facts as a distinct group of reproduction methods, as they have very specific features; see Chapter 9 for a more detailed discussion of recovery procedures.

Interiorization is less frequently discussed. For our purposes, we mention only that it includes visualization of abstract concepts; transformation of formal conventions into psychologically acceptable "rules of the game"; development of subconscious "parsing rules" for processing strings of symbols (most importantly, for reading mathematical expressions). At a more mundane level, one cannot learn an advanced technique of symbolic manipulation without first polishing one's skills in more routine computations to the level of almost automatic perfection. Interiorization is more than understanding; to handle mathematical objects effectively, one has to imprint at least some of their functions at the *subconscious* level of one's mind.

Interiorization is more than understanding; to handle mathematical objects effectively, one has to imprint at least some of their functions at the subconscious *level of one's mind.*

My use of the term "interiorization" is slightly different from the understanding of this word, say, by Weller et al. [142]. I put emphasis on the subconscious, neurophysiological components of the process. Meanwhile, I am happy to borrow from [142] the terms *encapsulation* (and the reverse procedure, *de-encapsulation*) to stand for the conversion of a mathematical procedure, a learned sequence of action, into an object. The processes of encapsulation and de-encapsulation are one of the principal themes of the book; see Section 6.1 for a more detailed discussion.

It is a popular misconception that mathematics is a dull repetitive activity.

It has to be clarified that reproduction does not mean *repetition*. It is a popular misconception that mathematics is a dull repetitive activity. Actually, mathematicians are easily bored by repetition. Perhaps this could create some difficulty in neurological studies of mathematics. Certain techniques for study of patterns of activation of the brain are easier to implement when the subject is engaged in an activity which is relatively simple and can be repeated again and again, so that the data can be averaged and errors of measurements suppressed. This works in studies like [203] which compared activation of the brains of amateur and professional musicians during actual or imagined performance of a short piece of violin music. Indeed, you can ask a musician to play the same several bars of music 10, 20, perhaps even 100 times—this is what they do in rehearsals. But it is impossible to repeat the same calculation 20 times: very soon the subject will remember the final and intermediate results. Moreover, most mathematicians will treat as an insult a request to repeat a similar calculation 20 times with varying data.

Proof is the key ingredient of the emotional side of mathematics.

Some mathematical activities are of a compound nature and can be used as means of both interiorization and reproduction. A really remarkable one is the generation of examples, especially very simple (ideally, the simplest possible) examples—as discussed in Section 1.1. Really useful examples can be loosely divided into two groups: “typical”, generic examples of the theory; or “simplest possible”, almost degenerate examples, which emphasize the limitations and the logical structure of the theory. Of course, one of the attractive features of Coxeter Theory is that it is saturated by beautiful examples of both kinds; I discuss some “simplest” cases in Section 2.6.

Proof, being the highest level of reproduction activity, has an important interiorization aspect: as Yuri Manin stresses in his book *Provable and Unprovable*, a proof becomes such only after it is *accepted* (as the result of a highly rigorous process) [383, pp. 53–54]. Manin describes the act of acceptance as a social act; however, the importance of its personal, psychological component can hardly be overestimated. One also should note that proof is the key ingredient of the emotional side of mathematics; proof is the ultimate explanation of *why* something is true, and a good proof often has a powerful emotional impact, boosting confidence and encouraging further questions "why?".

Visualization is one of the most powerful techniques for interiorization. It anchors mathematical concepts and ideas firmly into one of the most powerful parts of our brain, the visual processing module. Returning to the principal example of this book, mirrors and reflections, I want to point out that finite reflection groups allow an approach to their study based on a systematic reduction of this whole range of complex geometric configurations to simple two- and three-dimensional special cases. Mathematically this is expressed by a theorem:

a finite reflection group is a Coxeter group.

Satyan L. Devadoss, aged 3. © Satyan L. Devadoss

Avoiding a technical discussion, this means, in particular, that all relations between elements in a reflection group are consequences of relations between *pairs* of generating reflections. But a pair of mirrors in the n-dimensional Euclidean space is no more sophisticated a configuration than a pair of lines on the plane, and all the properties of the former can be deduced from that of the latter. *This provides a mathematical explanation of why visualization is such an effective tool in the theory of finite reflection groups.*

2.5 How to draw an icosahedron on a blackboard

My understanding of visualization as an interiorization technique leads me to believe that drawing pictures, and devising new kinds of pictures to draw, is an important way of facilitating mathematical work. This means that pictures have to be treated as mathematical objects and, consequently, must be *reproducible*. Students in the classroom should be able to *draw right away* the figures we put on the blackboard.

I have to emphasize the difference between *drawings* or *sketches* which are supposed to be reproduced by the reader or student and more technically sophisticated illustrative material (I will call these *illustrations*), especially computer-generated images designed for the visualization of complex mathematical objects (see a book by Bill Casselman [327] for an introduction into the art of illustrating mathematical texts). It would be foolish to impose any restrictions on the technical perfection of illustrations. However, one should be

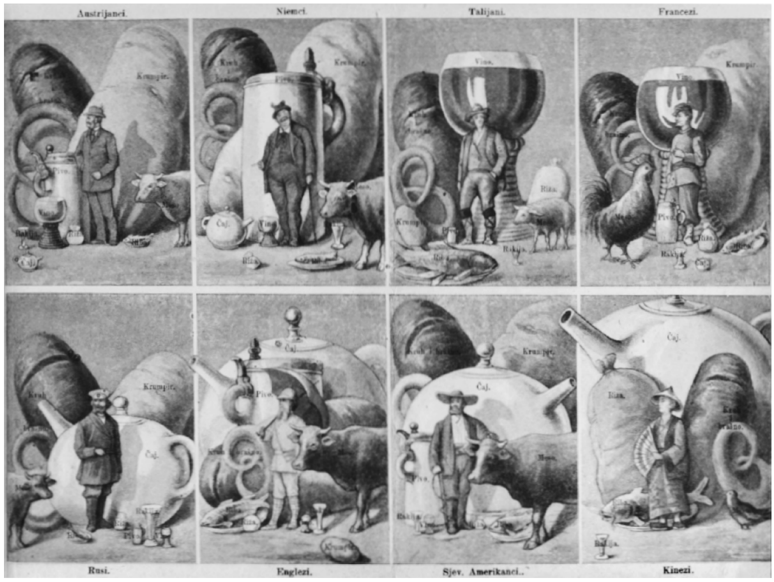


Fig. 2.3. What different nations eat and drink. A statistical diagram from a calendar published in Austro-Hungaria in 1901. Source: Marija Dalbello [19], reproduced with permission. See [20] for a discussion of the historical context.

This style of graphical representation of quantitative information strikes us now as patronizing and non-mathematical. It would be interesting to trace the cultural change over the 20th century: why do we expect a much more slim and abstract mode of presentation of information? Is this a result of the visual information overload created by TV and the Internet? It is worth mentioning that the level of basic numeracy in the middle classes of the Austro-Hungarian Empire, the target readership of the *Šareni svjetski koledar*, was almost definitely higher than in modern society.

aware of the danger of excessive details; as William Thurston—one of the leading geometers of our time—stresses,

> words, logic and detailed pictures rattling around can inhibit intuition and associations. [94, p. 165]

For that reason I believe that *drawings* should be intentionally very simple, even primitive. Mathematical pictures represent *mental* objects, not the real world! In the words of William Thurston,

> [people] do not have a very good built-in facility for *inverse vision*, that is, turning an internal spatial understanding back into a two-dimensional image. Consequently, mathe-

maticians usually have fewer and poorer figures in their papers and books than in their heads. [94, p. 164]

We have to be careful with our drawings and make sure that they correctly represent our "internal spatial understanding".

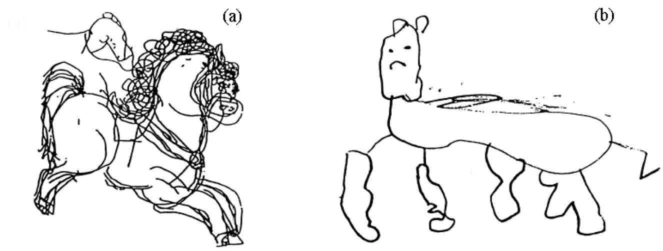


Fig. 2.4. Vision vs. "inverse vision": (a) a picture by Nadia (Drawing 3 from Selfe [228], © 1977 Elsevier, reproduced with permission) as opposed to (b) a picture by a normal child (Snyder and Mitchell [231], reproduced with permission). See [231] for a detailed discussion.

The pictures in Figure 2.4, taken from Selfe [228] and Snyder and Mitchell [231], illustrate the concept of "inverse vision" as introduced by Thurston. The picture (a) on the left is drawn from memory by Nadia, a three-and-a-half year old autistic child who at the time of making the picture has not yet developed speech [228]. Picture (b) is a representative drawing of a normal child, at age four years and two months. It is obvious that a normal child draws not a horse, but a concept of a horse.

Fig. 2.5. Horses by Nadia, at age of 3 years 5 months (Drawing 13 from Selfe [228], © 1977 Elsevier, reproduced with permission).

Fig. 2.6. Horses from Chauvet Cave (Ardeche). Document elaborated with the support of the French Ministry of Culture and Communication, Regional Direction for Cultural Affairs—Rhône-Alpes, Regional Department of Archaeology.

Nicholas Humphrey [195] drew even bolder conclusions from Nadia's miraculous drawings. He observed that Nadia's pictures have a most suggestive resemblance to cave paintings of 30,000–20,000 years ago—compare Figures 2.5 and 2.6. Humphrey conjectured that human language developed in two stages. At the first stage it referred only to people and relations between people; the natural world (including animals) had no symbolic representations in the language and therefore early people had no symbols for the external world. Cave paintings such as the one in Figure 2.6, are symbolically unprocessed images on the retina of the painter's eye, placed one over another without much coordination or a coherent plan. At the same time, people could already have words and symbols which referred to other people—which is consistent with the simultaneous presence, in some cave paintings, of strikingly realistic animals and highly schematic human figures; see Figure 2.7.

Mathematical pictures are symbolic images, not representations of reality.

Mathematical pictures are symbolic images, not representations of reality. Like a matchstick human in Figure 2.7, they are produced by "inverse vision". I dare to say that they do not belong to art. I propose that image processing which leads to the creation of paintings and drawings in the visual arts is different from that of mathematics.[8] Mathematical pictures therefore should not provoke an inferiority complex in readers who have not tried to draw anything since their days in ele-

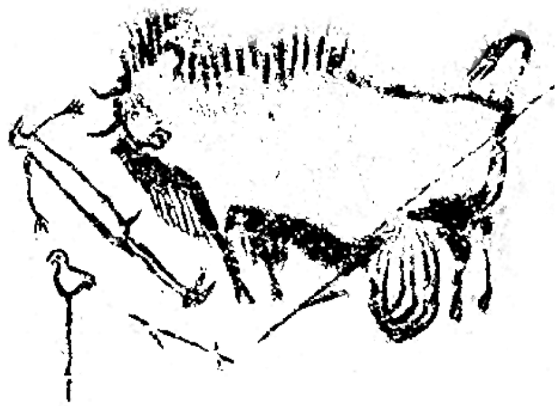


Fig. 2.7. A symbolic human and a naturalistic bull. Rock painting of a hunting scene, c. 17000 BC/ Caves of Lascaux, Dordogne, France. Source: *Wikipedia Commons*. Public domain.

mentary school; they should instead act as an invitation to readers to express their own mental images.

Figure 2.8 illustrates the most effective way of drawing an icosahedron, so simple that it is accessible to the reader with very modest drawing skills. First we mark symmetrically positioned segments in an alternating fashion on the faces of the cube (left) and then connect the endpoints (right). The drawing actually provides a proof of the existence of the icosahedron: varying the lengths of segments on the left cube, it is easy to see from continuity principles that, at a certain length of the segments, all edges of the inscribed polyhedron on the right become equal. [?] Moreover, this construction helps to prove that the group of symmetries of the resulting icosahedron is as big as one would expect it to be; see [294] for more details.

In a unit cube, find the length of the segments which makes all triangle faces equilateral; Figure 2.8.

Figure 2.8 works as a proof because it is produced by "inverse vision". To draw it, you have to run, in your head, the procedure for the construction of the icosahedron. And, of course, the continuity principles used are self-evident—they are part of the same mechanisms of perception of motion which glue, in our minds, the cinema's 24 frames per second into continuous motion.

I hope that now you will agree that Figure 2.8 deserves to be treated as a mathematical statement. It is useful to place it in a wider context. Notice that construction of the icosahedron is the same thing as construction of the finite reflection group H_3; this can be done by means of linear algebra—which leads to rather nasty calculations—or by means of representation theory—which

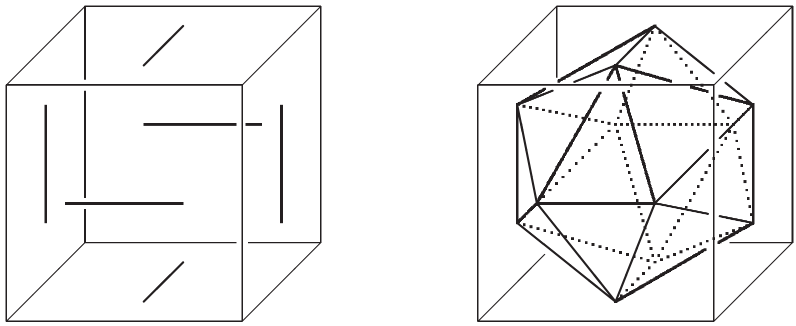


Fig. 2.8. A self-evident construction of an icosahedron. Drawing by Anna Borovik.

This construction of the icosahedron is adapted from the method of H. M. Taylor [47, pp. 491–492]. John Stillwell has kindly pointed out that it goes back to Piero della Francesca and can be found in his unpublished manuscript *Libellus de quinque corporibus regolaribus* from around 1480.

requires some knowledge of representation theory. It also can be done by quaternions—which is nice and beautiful but requires knowledge of quaternions. The graphical construction is the simplest; using computer jargon, it is a WYSIWYG ("What You See Is What You Get") mode of doing mathematics, which deserves to be used at every opportunity.

2.6 Self-explanatory diagrams

This section is more technical and can be skipped.

Self-explanatory diagrams have been virtually expunged from modern mathematics. I believe they can be useful, not only in proofs, etc., but also as the means of a metamathematical discussion of the structure and interrelations of mathematical theories.

Figure 2.9 is one example, taken from *Mirrors and Reflections* [294]: the isomorphism of the root systems D_3 (shown on the left, inscribed into the unit cube $[-1,1]^3$) and A_3 is not immediately obvious, but the corresponding mirror systems coincide most obviously. The mirror system D_3 (the system of mirrors of symmetry of the cube) is shown in the middle by tracing the intersections of mirrors with the surface of the cube and, on the right, by intersections with the surface of the tetrahedron inscribed in the cube. Comparing the last two pictures, we see that the mirror system of type D_3 is isomorphic to the mirror system of the regular tetrahedron, that is, to the system of type A_3.

As we shall soon see, this isomorphism has far-reaching implications.

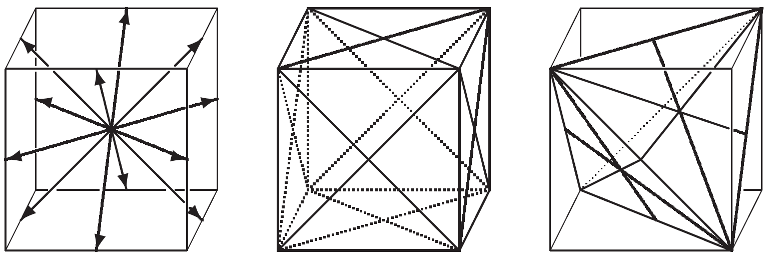


Fig. 2.9. An example of a self-explanatory diagram. Drawing by Anna Borovik.

Indeed, at the level of complex Lie groups the isomorphism $D_3 \simeq A_3$ becomes a rather mysterious isomorphism between the six-dimensional orthogonal group $\mathrm{SO}_6(\mathbb{C})$ and $\frac{1}{2}\mathrm{SL}_4(\mathbb{C})$, the factor group of the four-dimensional special linear group $\mathrm{SL}_4(\mathbb{C})$ by the group of scalar matrices with diagonal entries ± 1 (or, if you prefer to work with spinor groups, between $\mathrm{Spin}_6(\mathbb{C})$ and $\mathrm{SL}_4(\mathbb{C})$).

This is not yet the end of the story. The compact form of $\mathrm{SL}_4(\mathbb{C})$ is SU_4, and hence the embedding

$$\mathrm{SU}_4 \hookrightarrow \mathrm{Spin}_6(\mathbb{C})$$

features prominently in the representation theory of SU_4, and hence in the SU_4-symmetry formalism of theoretical physics.

But the underlying reason for the isomorphisms retains all the audacity of Keplerian reductionism: the tetrahedron can be inscribed into the cube. Compare with Figure 2.10.

Because of their truly fundamental role in mathematics, even the simplest diagrams concerning finite reflection groups (or finite mirror systems, or root systems—the languages are equivalent) have interpretations of cosmological proportions. Figure 2.11 is even more instructive. It is a classical case of the *simplest possible example* as discussed in Chapter 1. For example, it is the simplest rank 2 root system, or the simplest root system with a non-trivial graph automorphism; the latter, as we shall see in a minute, has really significant implications.

Figure 2.11 also demonstrates that the root system $D_2 = \{\pm\epsilon_1 \pm \epsilon_2\}$ is isomorphic to $A_1 \oplus A_1 = \{\pm\epsilon_1, \pm\epsilon_2\}$. At the level of Lie groups, this isomorphism plays an important role in the description of the structure of four-dimensional space-time of special relativity; namely, it yields the structure of the Minkowski group (the group of isometries of the four-dimensional space-time of spe-

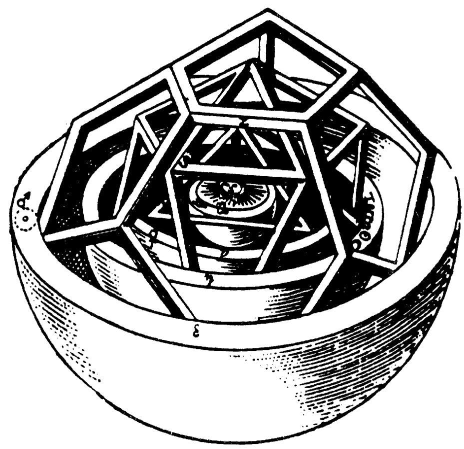


Fig. 2.10. A fragment of a famous engraving from Kepler's *Mysterium Cosmographicum*. Public domain.

cial relativity theory with the metric given by the quadratic form $x^2 + y^2 + z^2 - t^2$).

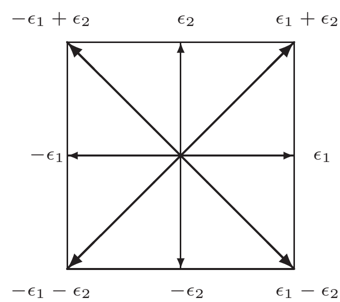


Fig. 2.11. This diagram demonstrates the isomorphism of the root systems $D_2 = \{\, \pm\epsilon_1 \pm \epsilon_2 \,\}$ and $A_1 \oplus A_1 = \{\, \pm\epsilon_1, \pm\epsilon_2 \,\}$. Drawing by Anna Borovik.

Indeed, the isomorphism of root systems $D_2 \simeq A_1 \oplus A_1$ leads to the isomorphisms

$$\mathrm{Spin}_4(\mathbb{C}) \simeq \mathrm{SL}_2(\mathbb{C}) \times \mathrm{SL}_2(\mathbb{C})$$

and

$$\mathrm{SO}_4(\mathbb{C}) \simeq \mathrm{SL}_2(\mathbb{C}) \otimes \mathrm{SL}_2(\mathbb{C})$$

(the tensor product of two copies of $\mathrm{SL}_2(\mathbb{C})$, each acting on its canonical two-dimensional space $\mathbb{C}^2$). The connected component of the Minkowski group is a real form of $\mathrm{SO}_4(\mathbb{C})$. Hence it is the group of fixed points of some involutory automorphism τ of the group $\mathrm{SO}_4(\mathbb{C})$. What is this automorphism τ? Let us look again at the quadratic form $x^2 + y^2 + z^2 - t^2$; it is a real form of the complex quadratic form $z_1^2 + z_2^2 + z_3^2 + z_4^2$ but has lost the symmetric pattern of coefficients. One can see that this means that τ swaps the two copies of $\mathrm{SL}_2(\mathbb{C})$ in $\mathrm{SL}_2(\mathbb{C}) \otimes \mathrm{SL}_2(\mathbb{C})$ and therefore has to be the symmetry between the two diagonals of the square in Figure 2.11. Being an involution, τ fixes pointwise the "diagonal" subgroup in $\mathrm{SL}_2(\mathbb{C}) \otimes \mathrm{SL}_2(\mathbb{C})$ isomorphic to $\mathrm{PSL}_2(\mathbb{C})$. (It is $\mathrm{PSL}_2(\mathbb{C})$ rather than $\mathrm{SL}_2(\mathbb{C})$ because its center $\langle -\mathrm{Id} \otimes -\mathrm{Id} \rangle$ is killed in the tensor product.) Hence the connected component of the Minkowski group is isomorphic to $\mathrm{PSL}_2(\mathbb{C})$.

Three cheers for Kepler!

Notes

[1] *Groupes et Algebras de Lie, Chap 4, 5, et 6* is one of the better books by Bourbaki; it even contains a drawing, in an unexpected deviation from his usual aesthetics. See an instructive discussion of the history of this volume by its main contributor, Pierre Cartier [87].

[2] See discussion of electromagnetic imaging in fish in Nelson [209], Rasnow and Bower [221].

[3] To reproduce Tarr's experiments, I was using PAINTSHOP PRO, with three-dimensional images produced by XARA, two software packages picked up from the cover CD of a computer magazine.

[4] PICTORIAL PROOFS. Jody Azzouni [5, p. 125] commented on pictorial proofs that they work only because we impose many assumptions on diagrams admissible as part of such proofs. As he put it:

> We can conveniently stipulate the properties of *circles* and take them as mechanically recognizable because there are no *ellipses* (for example) in the system. Introduce (arbitrary) *ellipses* and it becomes impossible to tell whether what we have drawn in front of us is a *circle* or an *ellipse*.

It is likely that his remark would not surprise cognitive psychologists; they believe that this is what our brains are doing anyway.

[5] FOUR-DIMENSIONAL INTUITION. Here is an interesting question: can one be habituated in a four-dimensional space (say, with a flight simulator). Of course, a three-dimensional image, stereo or holographic, could help. To put the point more radically, do we learn the number of dimensions?

[6] EULER'S THEOREM. I accept that the reader has every right to insist, if so inclined, that the "best" way to prove Euler's Theorem is by reduction to algebra: the characteristic polynomial of a "generic" three-dimensional orthogonal matrix is a cubic with real coefficients, hence has a real root and a pair of conjugate complex roots; the orthogonality means that the eigenvalues have magnitude 1, hence should be equal to ± 1 and $\cos\theta \pm i \sin\theta$.

If the matrix has determinant $+1$, then the real eigenvalue is $+1$, and the corresponding eigenvector gives the direction of the axis of rotation, while θ is the angle of rotation. But is that really better than Coxeter's proof?

[7]WHY WAS GEOMETRIC INTUITION EXPELLED FROM CLASSROOM? One of the reasons is the predominance of written examinations. There are no intrinsic pedagogical reasons why oral examinations are inferior to written examinations: written assessment dominates the modern teaching because it creates a convenient audit trail. In a written examination, purely "algebraic" solutions are easier to write down and easier to mark. In an oral exam, a candidate's awkward sketch on a blackboard may be worth a thousand words.

As a result of the degradation of academic practice under the pressure of the audit culture, I have seen courses in linear programming taught without any reference to the geometric interpretation of its principal concepts.

[8]The situation could be different in ornamental art, especially when images of other people and of the natural world are prohibited by cultural conventions or religion. The creators of the Islamic mosaics in the Alhambra had in fact discovered most of the planar crystallographic groups—an intellectual achievement which firmly places their work in the realm of mathematics. See Branko Grünbaum [42] for one of the most up-to-date discussion of symmetry groups present in the Alhambra. Also, Lu and Steinhardt [66] discuss an even more striking discovery of medieval Islamic architecture: quasi-periodic tilings.

3

The Wing of the Hummingbird

3.1 Parsing

So far I have emphasized the role of visualization in mathematics and its power of persuasion. Here I will try to unite the visual and symbolic aspects of mathematics and touch upon the limitations of visualization.

Indeed, visualization works perfectly well in the geometric theory of *finite* reflection groups, but it needs to be refined for the more general theory of infinite Coxeter groups. We take a brief look at this more general theory, which is of special interest to us at this point. As truly fundamental mathematical objects, Coxeter groups provide an example of a theory where the links between mathematical teaching and learning and cognitive psychology lie exposed. Besides the power of geometric interpretation and visualization, the theory of Coxeter groups relies on manipulation of *words* in canonical generators (chains of consecutive reflections, in the case of reflection groups) and provides one of the best examples of the effectiveness of the *language metaphor* in mathematics.

It is tempting to try to link the psychology of symbolic manipulation in mathematics with the Chomskian conjecture that humans have an innate facility for parsing human language. Basically, parsing is the recognition/identification of the structure of a string of symbols (phonemes, letters, etc.). We parse everything we read or hear. Here is an example from Steven Pinker's book [219, pp. 203–205] where this thesis is vigorously promoted:

> Remarkable is the rapidity of the motion of the wing of the hummingbird.

To make sense of the phrase, we have to mentally bracket subphrases, resulting in something like the following:

[Remarkable is
[the rapidity of
[the motion of
[the wing of
[the hummingbird]]]]].

A sentence might have a different bracket pattern. Just compare

[Remarkable is [the rapidity of [the motion]]]

and

[[The rapidity [that [the motion] has]] is remarkable].

Some patterns are harder to deal with than others: for example,

[[The rapidity that [the motion that [the wing] has] has] is remarkable].

Gregory Cherlin kindly offered a brainteaser from his childhood: Punctuate: Smith where Jones had had had had had had had had had had had the professor's approval.

Some bracketings are close to incomprehensible, even though the sentence conveys the same message:

[[The rapidity that [the motion that [the wing that [the hummingbird] has] has] has] is remarkable]. [?] [1]

Different human languages have different grammars, resulting in different parsing patterns. The grammar is not innate; Pinker emphasizes that what is innate is the human capacity to generate parsing rules. The generation of parsing patterns is a part of language learning (and young children are extremely efficient at it). It is also a part of the interiorization of mental objects of mathematics, especially when these *objects* are represented by strings of *symbols*.[2]

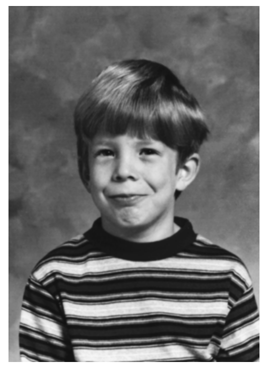
Cognitive scientists are very much attracted to case studies of "savants", autistic persons with an ability to handle arithmetic or calendrical calculations disproportionate to their low general IQ. As Snyder and Mitchell formulated it [231],

> ... savant skills for integer arithmetic ... arise from an ability to access some mental process which is common to us all, but which is not readily accessible to normal individuals.

The parsing mechanisms of the human brain are the key to the understanding of low-level arithmetic and formula processing.

What are these "hidden" processes? In one of the extreme cases (mentioned by Butterworth [160]), a severely autistic young man was unable to understand speech, but he could handle factors and primes in numbers. This suggests that certain mathematical actions are related not so much to language itself, but to the parsing facility, one of the components of the language system. An autistic person may have difficulty in handling language for reasons unrelated to his parsing ability; for example he may fail to recognize the source of speech communication as another person (or to understand the difference between what he knows and what the other person knows). But, in order to achieve such feats as "doubling 8 388 628 up to 24 times to obtain 140 737 488 355 328 in several seconds" [231, p. 589], an autistic person still has to be able to input into his brain the numbers given, inevitably, as strings of phonemes or digits.

I propose a conjecture that the parsing mechanisms of the human brain are the key to the understanding of low-level arithmetic and formula processing.



David Pierce, aged 6

Moving several levels up the hierarchy of mathematical processes, we have a fascinating idea in the theory of automatic theorem proving: *rippling*, a formalization of a common way of mathematical reasoning where "formulae are manipulated in a way that increases their similarities by incrementally decreasing their differences" [325, p. 13]. This is facilitated by subdividing the formula into parts which have to be preserved and parts which have to be changed. Again, we see that in order to understand how humans use rippling in mathematical thinking (and whether they actually use it), we have to understand how our brains parse mathematical formulae.

To be on the cautious side, I am prepared to accept that parsing might be much more prominent in the input/output functions of the brain than in the internal processing of information. In a rare case of a savant with higher than normal general intellectual abilities, Daniel Tammet is able to vividly describe the way he perceives the world, language, and numbers. It is obvious from his words that number processing happens to be directly wired into the visual module of his brain. For him, many numbers have a unique visual form.

> "Different numbers have different colours, shapes and textures . . . [The number] one is very bright and shining, like someone flashing a light into my face. Two is like a movement from right to left. Five is a clap of thunder or the sound of a wave against a rock. Six I find more difficult: it's more like a hole or a chasm. When I multiply numbers, I see two shapes in a landscape. The space between the images makes a third shape, like a jigsaw piece. And that third shape gradually crystallises: I see a fuzziness that becomes clearer and clearer." [430]

He adds that the whole process takes place in a flash, "like sparks flying off".

Although Daniel Tammet suffers from Asperger's syndrome (a form of autism) which to some degree inhibits his social skills—he has to remind himself that other people have thoughts entirely separate from his own and not to assume that they automatically know everything he knows—he has outstanding linguistic skills, speaks seven languages, and learned Icelandic in a week. He can also recite π to 22,514 decimal places. His case appears to confirm the thesis by Snyder and Mitchell; indeed, he has "an ability to access some mental process ... which is not readily accessible to normal individuals". This very access, however, requires parsing of the input.

3.2 Number sense and grammar

I turn to another remarkable insight from cognitive psychology, which links mechanisms of language processing to mastering arithmetic.

When infants learn to speak (in English) and count, there is a distinctive period, lasting five to six months, in their development, when they know the words *one, two, three, four* but can correctly apply only the numeral "one", when talking about a single object; they apply the words "two, three, four", apparently at random, to any collection of more than one object. Susan Carey [162] calls the children at this stage *one-knowers*. The most natural explanation is that they react to the formal grammatical structures of the adults' speech: *one doll*, but *two doll**S**, three doll**S***. At the next stage of development, they suddenly start using the numerals *two, three, four, five* correctly. Chinese and Japanese children become one-knowers a few months later—because the grammar of their languages has no specific markers for singular or plural in nouns, verbs, and adjectives.

In learning basic arithmetic, grammar precedes the words!

When the native language is Russian, the "one-knower" stage is replaced by the "one-(two-three-four) knower" stage, where children differentiate between three categories of quantities: single object sets, the sets of two, three, or four objects (without further differentiation between, say, two or three objects), and sets with five or more objects. This happens because morphological differentiation of plural forms goes further in Russian than in English.

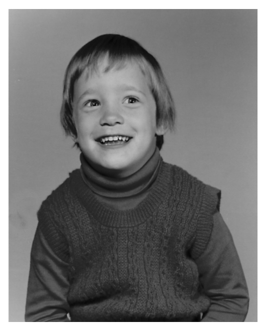
When I heard about special plural forms of two, three, or four nouns in a lecture by Susan Carey at the *Mathematical Knowl-*

edge 2004 conference in Cambridge, I was mildly amused because it made no sense to me, as a native Russian speaker. Still, I started to write on note paper:

one doll	одна кукл**А**
two doll**S**	две кукл**Ы**
three doll**S**	три кукл**Ы**
four doll**S**	четыре кукл**Ы**
five doll**S**	пять кук**ОЛ**
⋮	⋮
ten doll**S**	десять кук**ОЛ**

I was startled: yes, Susan Carey was right! I had been using, all my life, the morphological rules for forming plurals—but using them subconsciously, without ever paying attention to them. But, apparently, an infant's brain is tuned exactly to picking up the rules: it is easier for the child to associate the number of objects with the morphological marker in the noun signifying the object than with the words *one* or *two*. The interested reader will find a detailed discussion of plurality marking in Sarnecka et al. [225]. Meanwhile, one of the readers of my blog brought my attention to an even more striking example: in Russian, in some rare cases, the whole noun changes, not just the plurality marker. For example, one year, two year**S**, three, four, five year**S** are translated into Russian as один год, два, три, четыре год**А**, пять **ЛЕТ**. Notice the same thresholds: one/two, four/five. *In learning basic arithmetic, grammar precedes the words!*

We shall return to the discussion of the four/five threshold in the context of subitizing and short-term memory, in Section 4.1. However, in the particular case of the Russian language, there is the possibility of a historic explanation for the peculiar behavior of plurality markers: they are remnants from the times when an Indo-European predecessor of the Russian language used a system of numerals based on the number 4 [438]. It might happen, however, that the historic explanation is only intermediate, since it does not answer the crucial question of why a base 4 system had appeared in the first instance and why, apparently, its subsequent evolution led to a bifurcation into a base 9 system of numerals (now extinct, but still traceable in formulae from Russian fairy tales: в тридевятом царстве, *in a three times ninth kingdom*) and the decimal one, now predominant.[3]



Barbara Sarnecka, aged 3

3.3 What about music?

It would be interesting to see to what extent the parsing mechanisms of language processing are at work in wider auditory perception; for example, are they relevant for the perception of music? Do we parse notes by the same neurological mechanism which we use for parsing phonemes? Unfortunately, I cannot regard myself as an expert in music and therefore restrict myself to a few quotations.

My first quotation comes from a review, written by composer Dorothy Kerr, of the recent book *Music and Mathematics* [32] (strongly recommended!). Kerr, in effect, links music with the predictive nature of auditory processing.

> For a composer, some of the moments of greatest excitement lie in achieving a successful integration of 'mathematical' and 'musical' processes, though we may not think about it in these terms. Take the canon (a musical device that is essentially a translational symmetry) as an example: a very simple experiment that anyone can do is to set up a time delay between two copies of the same sound source (such as that produced when listening to digital radio simultaneously with an analogue receiver).[4] At first—provided the time interval allows it to be readily perceived—this simple geometrical effect can be very engaging to the ear (given how easy it is to create a satisfying effect in this way it is perhaps not surprising that canon is one of the earliest and most prevalent devices of musical composition). After the canon we have made has been going for a while, the novelty wears off and we develop the need for some kind of change or a new layer of interest. The nature and precise timing of such alterations, a calculation we usually make using our intuition, is one of the most basic aspects of the art of composition. [. . .] A process that is too obvious trails far behind the listener's ability to predict its outcomes. (Such music—to borrow the words of Harrison Birtwistle—'finishes before it stops'.)

The second quotation is from Thomas Mann's *Der Zauberberg*, a book famous for—among other things—a detailed study of the phenomenology of time. It describes music as parsing in its purest form:

> "I am far from being particularly musical, and then the pieces they play are not exactly elevating, neither classic nor modern, but just the ordinary band-music. Still, it is a pleasant change. It takes up a couple of hours very decently; I mean it breaks them up and fills them in, so there is something to them, by comparison with the other days, hours

> and weeks that whisk by like nothing at all. You see an unpretentious concert-number lasts perhaps seven minutes, and those seven minutes amount to something; they have a beginning and an end, they stand out, they don't easily slip into the regular humdrum round and get lost. Besides they are again divided up by the figures of the piece that is being played, and these again into beats, so there is always something going on, and every moment has a certain meaning, something you can take hold of . . . " (Translation by H. T. Lowe-Porter)

Of course, this gives only one dimension of music, essentially ignoring the harmony. In the words of Daniel Barenboim,

> The music can only be of interest if the different strands of the polyphonic texture are played so distinctly that they can all be heard and create a three-dimensional effect—just as in painting, where something is moved into the foreground and something else into background, making one appear closer to the viewer than the other, although the painting is flat and one-dimensional.

I would not dare to venture further and I leave it to someone else to develop this wonderful theme.

3.4 Palindromes and mirrors

To illustrate the role of parsing and other word processing mechanisms in doing mathematics, let us briefly describe Coxeter groups in terms of words.

We work with an alphabet $\mathbb{A}$ consisting of finitely many letters, which we denote a, b, etc. A *word* is any finite sequence of letters, possibly empty (we denote the empty word ϵ). Notice that we have infinitely many words. To impose an algebraic structure onto the amorphous mass of words, we proclaim that some of them are equivalent to (or synonymous with) other words; we shall denote the equivalence of words V and W by writing $V \equiv W$. We demand that concatenation of words preserve equivalence: if $U \equiv V$, then $UW \equiv VW$ and $WU \equiv WV$: if *mail* is the same as *post*, then *mailroom* is the same as *postroom*. We denote the language defined by the equivalence relation $\equiv$ by $\mathcal{L}_\equiv$.

So far all that was just the proverbial "general nonsense" which we frequently find in the formal exposition of mathematical theories. Mathematicians treat such formalities with great respect but frequently ignore them in actual work; formal definitions play the same role as fine print in insurance policies. Beware the fine print when you make a claim!

It is remarkable how little we have to add in order to create the extremely rigid, crystalline structure of a Coxeter group. To that end, we say that a word is *reduced* if it is not equivalent to any shorter word. Now we introduce just two axioms which define *Coxeter languages*:

DELETION PROPERTY. If a word is *not* reduced, then it is equivalent to a word obtained from it by deleting some *two* letters. (Of course, it may happen that the new word is still not reduced, in which case the process continues in the same fashion, two letters at a time.)

REFLEXIVITY. Words like aa obtained by doubling a letter are not reduced (hence are equivalent to the empty word, by the Deletion Property); *aardvark* is not a reduced word.


Michel Las Vergnas, aged 9

Actually, a Coxeter language is exactly a Coxeter group, but I intentionally ignore this crucial (for a mathematician) fact and formulate everything in terms of words and languages.

I will now give a (straightforward) reformulation of a classical theorem of 20th-century algebra, due to Coxeter and Tits. My formulation is a bit of a caricature devised specifically for the purposes of the present book.

To emphasize the language aspects, let us make *palindromes*, that is, non-empty reduced words such as "level" that read the same backwards as forwards, the central object of the theory. [5]

Now the Coxeter–Tits Theorem becomes a theorem about *representation of palindromes by mirrors*.

The Palindrome Representation Theorem. Assume that a Coxeter language $\mathcal{L}_\equiv$ contains, up to equivalence, only finitely many palindromes.[6] Then:

- There exists a finite closed system $\mathbb{M}$ of mirrors in a finite-dimensional Euclidean space $\mathbb{R}^n$ such that the mirrors in $\mathbb{M}$ are in one-to-one correspondence with the equivalence classes of the palindromes.
- Moreover, if M_1 and M_2 are mirrors and P_1, P_2 their palindromes, then the palindrome associated with the reflected image of the mirror M_1 in the mirror M_2 is $P_2P_1P_2$, if the latter is reduced, or a palindrome obtained from the word $P_2P_1P_2$ by reduction.
- Finally, every closed finite system of mirrors in the Euclidean space $\mathbb{R}^n$ can be obtained in this way from the system of palindromes in an appropriate Coxeter language.

The interested reader will find the ingredients of a proof of this result in Chapters 5 and 7 of [320]. It involves, at some point, the

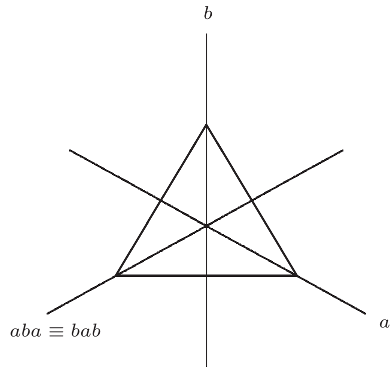


Fig. 3.1. The Palindrome Representation Theorem: The three mirrors of symmetry of the equilateral triangle correspond to the palindromes a, b, and aba. Together with the equivalences $aa \equiv bb \equiv \epsilon$ (the empty word), the equivalence $aba \equiv bab$ warrants that the corresponding Coxeter language does not contain any other palindromes.

following equivalence [294, Exercise 11.8]:

$$a_1 \cdots a_l \equiv a_l^{a_{l-1}\cdots a_1} \cdot a_{l-1}^{a_{l-2}\cdots a_1} \cdots a_2^{a_1} \cdot a_1,$$

where $b^{a_k \cdots a_1}$ is an abbreviation of a palindromic word

$$a_1 \cdots a_k \cdot b \cdot a_k \cdots a_1;$$

the foregoing identity expresses an arbitrary word as the concatenation of palindromic words; its proof consists of the rearrangement of brackets and the cancelation of doubled letters $a_i a_i$ whenever they appear. Proofs like that are one of the many reasons why, in order to master the theory of Coxeter groups expressed in a "linguistic" manner, the novice reader has to develop an ability to manipulate imaginary mental brackets with a rapidity comparable with only the remarkable rapidity of the motion of the wing of the hummingbird.

I reiterate that I devised the palindrome formulation of the Representation Theorem specifically for the needs of the present book. When afterwards I made a standard search on GOOGLE and MathSciNet [446], I was pleased to discover that my formulation appeared to be new.

We can reuse space, but, unfortunately, we cannot reuse time.

I was also pleasantly surprised to find more than a hundred papers on palindromes produced by computer scientists. The set of all palindromic words in a given alphabet is one of the simplest examples of a language which can be generated only by a device with some kind of memory, say, with a stack or push-down storage which works on the principle “last in—first out”, like bullets in a handgun clip. It makes palindromes a very attractive test problem in the study of the complexity of word processing, for example, for comparing two fundamental concepts of algorithmic complexity: space-complexity, measured by the amount of memory required, and time-complexity. The difference between the two complexities is deeply philosophical: we can reuse space, but, unfortunately, we cannot reuse time. I was particularly fascinated to learn that palindromes are recognizable by Turing machines working within sublogarithmic space constraints [409]. Hence, in this particular problem, it is possible to overwrite and reuse the memory.

Perhaps it is exactly the necessity to engage—and reuse—one’s low-level memory that turns palindromes into such popular and addictive brainteasers.

3.5 Parsing, continued: do brackets matter?

Understanding the role of interiorization and reproduction is crucial for any serious discussion of what is actually happening in teaching and learning mathematics, and it is very worrisome that this cognitive core is so frequently absent from the professional discourse on mathematical education. This is especially true for the discussion of the merits of computer-assisted learning of mathematics, where the use of technology has changed the cognitive content of standard elementary routines which for centuries served as building blocks for learning mathematics.

Typing a command is like saying a sentence, while clicking a mouse is equivalent to pointing a finger in conversation.

And here is a small case study. For some years I had been teaching courses in mathematical logic based on two well-known software packages: SYMLOG [305] and TARSKI’S WORLD [293] (reviews: [107, 115, 118]). SYMLOG used a DOS command line interface which was extremely weak even by the standards of its time, while TARSKI’S WORLD very successfully exploited the graphical user interfaces of Apple and Windows for the visualization of one of the key concepts of logic, a model for a set of formulae (see [8] for the discussion of the underlying philosophy and [403] for underlying mathematics—it is highly non-trivial). Also, TARSKI'S WORLD made a very clever use of games to explain another key concept, the validity of a formula in an interpretation (although the range of interpretations was limited [118]). However, when it came to a written test, students taught with SYMLOG made virtually no errors in the composition of logical formulae, while those taught with TARSKI'S WORLD very obviously struggled with this basic task. The reason was easy to find: SYMLOG's very unforgiving interface required retyping the whole formula if its syntax had not been recognized, while TARSKI'S WORLD's user-friendly formula editor automatically inserted matching brackets. Although TARSKI'S WORLD's students had no difficulty with rather tricky logic problems when they used a computer, their inability to handle formulae without a computer was alarming. Indeed, in mathematics, the ability to reproduce your mental work has to be media-independent. Relieving the students of a repetitive and seemingly mindless task led them to lose a chance to develop an essential skill.

It is appropriate to mention that, in parallel with visualization, there is another mode of interiorization, namely *verbalization*. Indeed, we understand and handle much better those processes and actions which we can describe in words. In naive terms, *typing a command is like saying a sentence, while clicking a mouse is equivalent to pointing a finger in conversation*. The reader would no doubt agree that, when teaching mathematics, we have to incite our students to speak. The tasks of opening and closing matching pairs of brackets, however dull and mundane they may be, activate deeply rooted neural mechanisms for the generation of parsing rules and are crucial for the interiorization of symbolic mathematical techniques.

Sadly, it appears to be acceptable to promote educational software without spelling out what students will lose *as a result of its use. In pharmaceutical research, a similar practice would constitute a criminal offence.*

I understand that my claims will inevitably provoke the stock response from the promoters of computer-assisted learning: computers are a valuable tool and they help students to save time wasted on routine calculations, allow them to concentrate on deeper conceptual understanding of mathematics, etc. I agree with all that. But I am concerned that the discourse on computer-assisted learning is anti-scientifically skewed and suffers from a cavalier approach to the assessment of the implications for the learner. In medical sciences, promotion of a new medicine without a careful study of its side effects is an academic, regulatory (and, frequently, criminal) offence. In educational circles, it appears to be acceptable to promote a new piece of software for learning a particular chapter of mathematics without spelling out what students will *lose* as a result of its use. Educational software has to be judged on the balance of gains and losses.

3.6 The mathematics of bracketing and Catalan numbers

> We have not begun to understand the relationship between combinatorics and conceptual mathematics.
> Jean Dieudonné [24]

The parsing examples we have considered so far have been of a special kind, *binary parenthesizing*; I do not want to venture into anything more sophisticated because even placing parentheses in an expression made by repeated use of a binary operation, such as

$$a + b + c + d,$$

is already an immensely rich mathematical procedure. In various disguises, it appears throughout all of mathematics. There is no better example than Richard Stanley's famous collection of 66 problems on Catalan numbers [406, Exercise 6.19, pp. 219–229] (solutions can be found in [407]). I mention a couple of examples.

The number of different ways to completely parenthesize the formal sum

$$a_0 + a_1 + \cdots + a_{n-1} + a_n \quad (n + 1 \text{ numbers})$$

is called the *n-th Catalan number* and is denoted C_n; it can be shown that

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

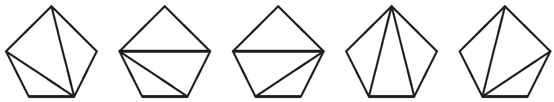
For example, when $n = 3$, we have 5 ways to place the brackets in $a + b + c + d$, namely:

$$a + (b + (c + d)), \quad a + ((b + c) + d), \quad (a + (b + c)) + d,$$

$$(a + b) + (c + d), \quad ((a + b) + c) + d$$

(following the usual convention, I skip the outermost pair of brackets).

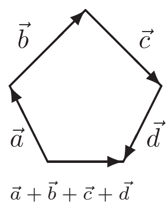
Remarkably, when you count ways to triangulate a convex $(n + 2)$-gon by $n - 1$ diagonals without crossing, you come to exactly the same result:



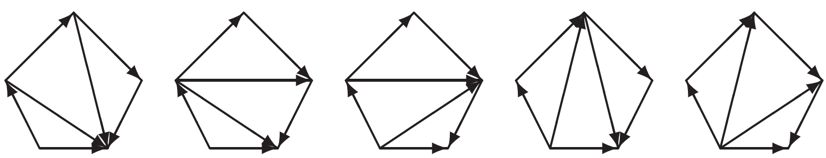
This mysterious coincidence is resolved as soon as we treat drawing diagonals as taking the sums of vectors

$$\vec{a} + \vec{b} + \vec{c} + \vec{d}$$

going along the $n+1$ sides of the $(n+2)$-gon, with the last side (the base of the polygon) representing the sum:



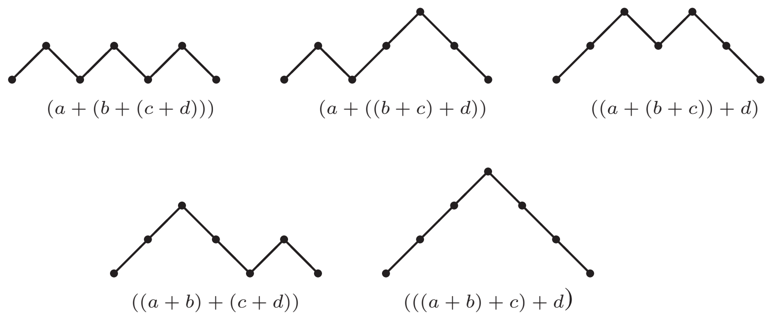
Now the one-to-one correspondence between parenthesizing the vector sum and drawing the diagonals becomes self-evident:



$\vec{a} + (\vec{b} + (\vec{c} + \vec{d}))$ $\vec{a} + ((\vec{b} + \vec{c}) + \vec{d})$ $(\vec{a} + (\vec{b} + \vec{c})) + \vec{d}$ $(\vec{a} + \vec{b}) + (\vec{c} + \vec{d})$ $((\vec{a} + \vec{b}) + \vec{c}) + \vec{d}$

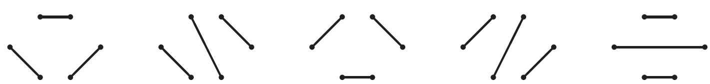
I do not remember the exact formulation of the problem which led me, as a schoolboy, to the discovery of this correspondence between parenthesizing and triangulations, but I remember my feeling of elation—it was awesome.

As a teaser to the reader I give another class of combinatorial objects which are also counted by Catalan numbers. Take graph paper with a square grid, and assume that the unit (smallest) squares have length 1. A *Dyck path* is a path in the grid with steps $(1,1)$ and $(1,-1)$. I claim that the number of Dyck paths from $(0,0)$ to $(2n,0)$ which never fall below the coordinate x-axis $y = 0$ is, again, the Catalan number C_n. I give here the list of such paths for $n = 3$, arranged in a natural one-to-one correspondence with the patterns of parentheses in $a + b + c + d$:



Can you describe the rule? Notice that I added, for your convenience, the exterior all-embracing pairs of parentheses; they are usually omitted in algebraic expressions. (Notice also that this correspondence gives, after some massaging, an algorithm for checking the formal correctness of bracketing—so that the algorithm says that the bracketing $(a + (b + c))$ is correct while $(a + b) + c) + ((d + e)$ are not.)

One more example is concerned with n non-intersecting chords joining $2n$ points on the circle:



Find a one-to-one correspondence between the 5 *chord diagrams and the* 5 *ways to parenthesize the sum* $a + b + c + d$*. Hint: there are* 3 *chords and* 3 *"+" symbols.*

Again, there are

$$C_n = \frac{1}{n+1} \cdot \binom{2n}{n}$$

different ways to draw the chords. [?]

Richard Stanley makes a wry comment on his list of Catalan number problems [406, pp. 219–229] that, ideally, the best way to solve all 66 problems is to construct directly the one-to-one correspondences between the 66 sets involved, $66 \cdot 65 = 4,290$ bijections in all! It is likely, however, that all 66 sets could be shown to be bijective to one specific set; the set of all rooted trivalent trees with n internal nodes is the most likely candidate for the special role since all 66 sets have a very distinctive hierarchical structure.

This is still not the end of the story: the striking influence of a seemingly mundane structure, grammatically correct parenthesizing, can be traced all the way back to the most sophisticated and advanced areas of modern mathematics research. A brief glance at one of Stasheff's associahedra (Figure 3.2) suggests that they live in the immediate vicinity of Coxeter Theory.[7] Actually, generalized associahedra can be defined for any finite Coxeter group (Stasheff's associahedra being associated, of course, with the symmetric group

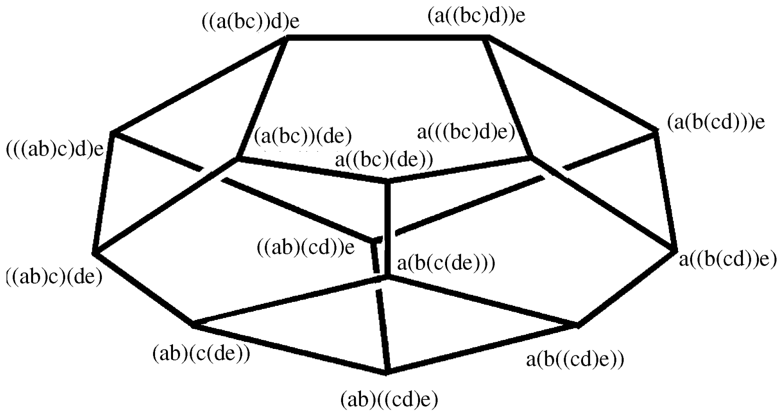


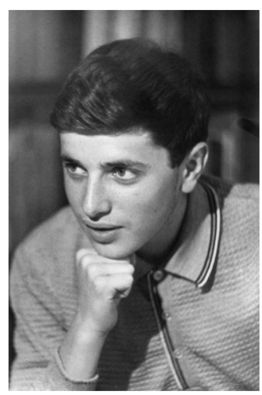
Fig. 3.2. Stasheff's associahedron: the binary parenthesizings of n symbols can be arranged as vertices of a convex $(n-2)$-gon, with two vertices connected by an edge if the corresponding parenthesizings differ by the position of just one pair of brackets.

Sym_n viewed as the Coxeter group of type A_{n-1}); for some recent results see, for example, Fomin and Zelevinsky [351].

3.7 The mystery of Hipparchus

It appears that the importance of parsing has been appreciated by mathematicians and philosophers since ancient times. The following fragment from Plutarch, a famous Greek biographer of the 2nd century A.D., remained a mystery for centuries:

> Chrysippus says that the number of compound propositions that can be made from only ten simple propositions exceeds a million. (Hipparchus, to be sure, refuted this by showing that on the affirmative side there are 103,049 compound statements, and on the negative side 310,952.)



Andrei Zelevinsky, aged 16

Here Plutarch refers to two prominent thinkers of Classical Greece: the philosopher Chrysippus (c. 280 B.C.–207 B.C.) and the astronomer Hipparchus (c. 190 B.C.–after 127 B.C.). Only in 1994 did David Hough notice that 103,049 is the number of arbitrary (non-binary) parenthesizings of 10 symbols, that is, the number of all possible expressions like

$$(xxxx)((xx)(xx)xx).$$

This numerical observation suggests that, for Chrysippus and Plutarch, “compound” propositions were built from “simple” propositions simply by bracketing.

The mathematics and history of Hipparchus’ number is discussed in detail in a paper by Richard Stanley [289]. The number of parenthesizings of n symbols is known as the *Schröder number* $s(n)$; the first 11 values of the Schröder numbers are

$$1,\ 1,\ 3,\ 11,\ 45,\ 197,\ 903,\ 4279,\ 20793,\ 103049,\ 518859.$$

In 1998, Laurent Habsieger, Maxim Kazarian, and Sergei Lando [269] suggested a very plausible explanation of the second Hipparchus number, of compound statements on the “negative side”. They observe that

$$\frac{s(10)+s(11)}{2} = 310,954$$

and, assuming a slight arithmetic or copying error in Plutarch’s text, suggest we interpret the compound statements on the “negative side” as parenthesizings of expressions

$$\text{NOT}\, x_1 x_2 \cdots x_{10}$$

under the following convention: the negation NOT is applied to all the simple propositions included in the first pair of brackets that includes NOT. This means that the parenthesizings

$$[\text{NOT}\, [P_1] \cdots [P_k]]$$

and

$$[\text{NOT}\, [[P_1] \cdots [P_k]]]$$

give the same result, and most of the negative compound propositions can be obtained in two different ways. The only case which is obtained in a unique way is when one only takes the negation of x_1. Therefore twice the number of negative compound propositions equals the total number of parenthesizings on a string of 11 elements

$$\text{NOT}\, x_1 x_2 \cdots x_{10}$$

plus the total number of parenthesizings on a string of 10 elements

$$(\text{NOT}\, x_1) x_2 \cdots x_{10}.$$

This, indeed, provides the value $(s(10)+s(11))/2 = 310,954$.

Nowadays, the thinkers of Classic Antiquity do not enjoy the same authority and revered status that they had up to the 19th century. Armed with the machinery of enumerative combinatorics, we may look condescendingly at the fantastic technical achievement of Hipparchus (which became possible perhaps only because he was an astronomer and could handle sophisticated arithmetic calculations, possibly using Babylonian base-60 arithmetic). But I find it highly significant that ancient Greek philosophers, in their quest for understanding of the logical structure of human thought, identified the problem of parsing and attempted to treat it mathematically.

Notes

[1]ONE MORE TEASER. Punctuate:

> Where a previous sentence had had had had had had had had had
> had had had had this sentence contains more.

Continue this inductively to give an arbitrarily large number of perfectly grammatically correct consecutive "hads" in the sentence. (Offered by a commentator on my blog who called himself Ben.)

[2]PARSING RULES. David Pierce has drawn my attention to an interesting question related to parsing rules for mathematical formulae. To what extent is the "infix" notation for binary operations and relations, when the symbol for operation or relation is placed between the symbols for objects, like $a + b$ and $a < b$, made more natural for humans by the nature of their innate grammar generating rules? Or is the predominance of infix notation more of a cultural phenomenon, a fossilized tradition? Why does "reverse Polish notation" (or "suffix" notation) puzzle most people when they first encounter it? In reverse Polish notation, the expression

$$(a + b) \times (c + d)$$

is written as

$$ab + cd + \times.$$

It has serious advantages in computing: when using a hand-held calculator designed and programmed for the use of reverse Polish notation, one is not troubled with saving the intermediate results into the memory; this is done automatically. On an ordinary calculator, one has to save the intermediate result $a + b$ when calculating $(a + b) \times (c + d)$. Notice that infix notation does not generalize from binary to ternary operations, and ternary operations and relations are not frequently found in mathematics. Is that because our writing is linear, reflecting the linear nature of speech? Words denoting ternary or higher arity relations are infrequent in human languages. The predicate "a is between c and d" is a noticeable exception in English. Interestingly, the "betweenness" relation among points on a line was famously absent from Euclid's axiomatization of geometry (see Section 11.4).

[3]SINGULAR, DUAL, TRIAL, PAUCAL. . . . In general, languages tend to treat numbers from 1 to 4 differently; see [196]. Owl remarked that traces of the dual category still can be found in Russian (and it is still present in

Slovenian. Apparently, dual was mostly purged from Russian in the language reform of Peter the Great.

Barbara Sarnecka wrote to me:

> . . . usually the options are
> (a) Singular (1)/Plural (2+),
> (b) Singular (1)/Dual (2)/Plural (3+),
> (c) Singular (1)/Dual (2)/Trial (3)/Plural (4+),
> (d) Singular (1)/Dual (2)/Paucal (approximately 3–4)/Plural (approximately 5+).
> Anyway, Russian is the only language I know of where the dual and paucal categories have been merged into one, so that is quite interesting. Is it possible that there was, earlier, a singular/dual/paucal/plural system, and that Peter [the Great] tried to simplify it by combining the dual and paucal categories?

I would be happy to learn more about pluralities—although this theme leads well beyond the scope of my book. In particular, I was intrigued by a comment from my Hungarian colleague that, in the Hungarian language, a plurality marker is present in nouns when no specific numeral is used with the noun.

As I have already mentioned in the main text, I shall return to the discussion of thresholds for pluralities in Section 4.1.

[4]CANON IN POETRY. My dear old friend Owl reminded me that canon can be found in poetry, where it is sometimes used with a totally mesmerizing effect:

В посаде, куда ни одна нога
Не ступала, лишь ворожеи да вьюги
Ступала нога, в бесноватой округе,
Где и то как убитые спят снега, –

Постой, **в посаде куда ни одна**
Нога не ступала, лишь ворожеи
Да вьюги ступала нога, до окна
Дохлестнулся обрывок шальной шлеи.
(Boris Pasternak)

[5] PALINDROMES AND COXETER GROUPS. My "palindrome" formulation of the Coxeter-Tits Theorem is one of many manifestations of a *cryptomorphism*, the remarkable capacity of mathematical concepts and facts for translation from one mathematical language to another; see more on that in Section 4.2. I recall again that, in this book, I have adopted a "local", "microscopic" viewpoint. Although the "palindrome theory" is of little "global" value for mathematics, it demonstrates some interesting "local" features of mathematics.

[6]Without the assumption about the finiteness of the number of palindromes, the Palindrome Representation Theorem is still true if we accept mirrors in non-Euclidean spaces. Section 5.1 contains some examples of mirror systems in the hyperbolic plane.

[7]An elementary construction of associahedra can be found in Loday [381].