ELSEVIER Linear Algebra and its Applications 302–303 (1999) 601–613

LINEAR ALGEBRA AND ITS APPLICATIONS

www.elsevier.com/locate/laa

Students' understanding of proofs: a historical analysis and implications for the teaching of geometry and linear algebra ☆

Guershon Harel *

Department of Mathematics, Purdue University, West Lafayette, IN 47907, USA

Received 10 December 1998; accepted 16 June 1999

Submitted by F. Uhlig

Dedicated to Hans Schneider

The process of observing and analyzing students' behaviors is interesting and complex but also unstable. It is unstable because it involves countless variables, many of which are uncontrollable. Despite this, what we learn from this process is useful, even essential, in designing and implementing mathematics curricula for both students and teachers. This presentation is about students' behaviors in relation to justification and proof. Some of these behaviors are assumed to be due to faulty instruction in school; others seem to be unavoidable, in the sense that they are of human developmental nature. Analyzed from a historical perspective of mathematical development, these students' understandings of proof can be classified into three categories:

- *Category* 1: In this category, students' understandings of proof (viewed in relation to those of their instructors) seem to parallel the Greek conception of mathematics (viewed in relation to that of modern days).
- *Category* 2: In this category, students' understandings of proof are reminiscent of the 16–17th century conception of mathematics.
- *Category* 3: In this category, students' understandings of proof seem, to a large extent, to be a result of faulty instruction in the elementary and secondary schools.

☆ This is the text of a plenary talk given at the International Linear Algebra Society Conference held in 1998 at the University of Wisconsin, Madison.

* E-mail: harel@math.purdue.edu

PII: S 0 0 2 4 - 3 7 9 5 (9 9) 0 0 1 3 9 - 1

In this presentation, I will focus on the first two categories, the third category is addressed in [1]. The findings about students' understandings of proof, part of which is reported here, were obtained from a sequence of six teaching experiments with a total of 169 students (mathematics majors and engineering majors). The data were collected from classroom observations in the form of field notes and retrospective notes, clinical interviews, homeworks, and written tests and quizzes. Some of the data came from videotaped classroom sessions and sixty-to-ninety-minute clinical interviews with students (for more detailed accounts, see [1]).

While the data from the experiments support the observation of students' difficulties with proof, the historical analysis I describe here only points to a *possible* parallelism between these difficulties and historical phenomena in the development of mathematics. My research did not intend to establish and has not established such a parallelism. The historical analysis has, however, provided me with insights as to the *possible* conceptual basis for students' mathematical behaviour.

1. Category 1: students' understanding of mathematics in relation to greek versus modern mathematical thought

The axiomatic conception of proof is when the student understands that at least in principle a mathematical justification must have started originally from undefined terms and axioms. There really are three different levels of the axiomatic conception, which historically were developed in three consecutive periods: *Intuitive axiomatic* conception, *structural* conception, and *axiomatizing* conception.

Intuitive axiomatic. The view of what constitutes an acceptable mathematical proof has had many turning points. Babylonian mathematics is considered proof-free, because it does not deal with general statements, deduction, or explanations; rather, it prescribes specific solutions to specific problems. The axiomatic method – that is, the notion of deductive proof from some accepted principles – was conceived by the Greeks. However, it is important to note that the Greeks had one single type of mental objects in mind, namely, objects that are idealizations of physical reality, such as a line, plane, triangle, etc. Accordingly, with the Intuitive Axiomatic conception, the student is able to handle only axioms that correspond to her/his intuition. For example, the statement "One and only one line goes through two points" is understood only in the context of personal geometric intuition. Here the objects, which are derived from an idealization of the physical reality, determine the set of axioms. Similarly, the statement, "For any a and b in F, $a + b = b + a$" is understood only in the context of experience with real numbers. In modern mathematics, on the other hand, the objects are determined by a set of axioms.

This was a revolutionary way of thinking in the development of mathematics and may shed light on some of the difficulties we observed with students.

Structural. To explain the second level, let me point to a critical distinction between Euclid's Elements and Hilbert's Grundlagen. While the Elements is restricted to a single interpretation – namely that its content is a presumed description of human spatial realization – the Grundlagen is open to different possible realizations, such as Euclidean space, the surface of a half-sphere, ordered pairs and triples of real numbers, etc. – including the interpretation that the axioms are meaningless formulas. In other words, the Grundlagen characterizes a STRUCTURE that fits different models. This obviously is not unique to geometry. In algebra, a group or a vector space is defined to be any system of objects satisfying certain axioms that specify the structure under consideration. Accordingly, the structural conception is the understanding that definitions and theorems represent situations from *different* realizations that share a common structure determined by a *permanent* set of axioms. In this conception, the axioms that define the structure are permanent, and one studies the structure itself, not just the axiom system. So, for example, one studies real analysis on the basis of the axioms of a complete ordered field, or one studies the theory of vector spaces on the basis of the vector space axioms, etc.

Axiomatizing. Our data suggest that the structural conception is a cognitive prerequisite to the *axiomatizing* proof conception – a conception by which a person is able to investigate the implications of varying a set of axioms, or to understand the idea of axiomatizing a certain field.

One of the questions we addressed in our studies was: Do undergraduate mathematics majors possess the axiomatic conception at any level? For example, do students understand that axioms in geometry require no specific interpretation? In particular, can students consider their own intuitive space (i.e., the Euclidean space) as a specific system that may or may not satisfy the structure at hand?

1.1. A sample of results

Textbooks in axiomatic geometry usually begin with finite geometries as a preparation for non-Euclidean geometries. As you see in Fig. 1 and Table 1, the idea that geometric properties are not supposed to evoke spatial imagery is virtually absent from students' conception. We should remember that this idea is a relatively new concept in mathematics; it was born at the turn of this century with the publication of Hilbert's Grundlagen. Poincare, in his review of the Grundlagen, saw a need to point to this seemingly self-evident feature of the Grundlagen. To our students this was a very difficult idea.

When students are unable to detach from a specific context, whether it is the context of intuitive Euclidean space in geometry or the context of $\mathbb{R}^n$ in linear algebra, we call that conception, *contextual*. And so, with the contextual

Problem: Consider a model where the points are A, B, C, D, E, F, G and the lines are {A,B,F}, {A,C,E}, {A,D,G},{B,C,D}, {B,E,G}, {C,F,G}, {D,E,F}. Do the properties below hold?

1. Only one line goes through two points
2. Given a line and a point not on it, there is a line containing the given point and is parallel to the given line.

Students' responses include statements like:

- "These are consecutive points."
- "The line continues in each direction."
- "It is a segment with three points."
- "A line with only 3 points? Impossible
- "The line has to be all the points contained between those points."
- "There has to be an infinite number of points between the points."

Fig. 1.

conception, general statements are interpreted (and proved) in terms of a specific context.

Judging form the historical development of geometry, we conjecture that the contextual conception is developmentally inevitable, but it should have been developed in the secondary school. What surprises us is its robust influence on students in an advanced stage in their mathematical education. We must appreciate, however, the non-triviality of the structural conception.

Here is another example from history that helps us appreciate this state of intellectual development. The modern notion of "number" was born when symbols representing no specific reality were treated as objects that can be operated upon by certain rules. These objects are defined not by what they represent but by an a priori set of rules. Not all mathematicians of the 17th century shared this new way of thinking; some raised serious doubts about its intelligibility and viability. How is it possible to reason about symbols without a concrete referent and especially without a geometrical referent, as in the case of imaginary numbers and negative numbers? How is it possible, asked Arnauld, a 17th century mathematician, to subtract a greater quantity from a smaller one, where the mental image of "quantity" is nothing else but a physical amount or a spatial capacity? Moreover, how is it possible to understand such a statement as $(-1)/1 = 1/(-1)$, where the quantity 1 is larger

Table 1

	N	%
Contextual (Imposition of extraneous notions, e.g., "betweeness", "distance", "collinearity"	17	55
Axiomatic	5	16
Undecided	9	29
Total	31	100

than the quantity -1, and therefore, the division of 1 by -1 must be smaller than the division of -1 by 1 [2]?

Let me conclude this category with two items: Examples of students' responses that demonstrate the absence of the structural conception from our students' reasoning (Fig. 2), and a quantitative comparison between the two conceptions in one of the teaching experiments (Table 2).

Students' inability to deal with any geometric structure but the one corresponding to their spatial imageries is reminiscent of the Greek's view of mathematics. In Greek science, concepts are formed in continual dependence on, and interpreted from the point of view of, their "natural" foundations, and their scientific meaning is abstracted from "natural," pre scientific experience. In modern science, on the other hand, what intended by the concept is not an object of immediate insight, but an object whose scientific meaning can be determined only by its connection to other concepts, by the total edifice to which it belongs, and by its function within this edifice [3].

To illustrate the phenomenon of how students are constrained by their physical imageries, consider the following example. When linear algebra instructors present to the students a problem such as "Given W is a subspace of $\mathbb{R}^n$, find the projection of c onto W" along with the sketch in Fig. 3, they do not mean the sketch to be literal but symbolic. It turned out that such a sketch is not conceived as a REPRESENTATION of the abstract setting, but as the ACTUAL OBJECT of inquiry.

Another illustration of students' appeal to their immediate physical reality rather than to the actual definition comes from their difficulties with the concepts of span, dependence, and independence. For example, the sketches in Fig. 4, when presented to illustrate the definition,

$$\mathrm{span}\{a_1,\ldots,a_k\} = \{x_1a_1 + \cdots x_ka_k | x_1,\ldots,x_k \text{ are real numbers}\},$$

are not conceived as a special cases of the concept of span, but the ACTUAL OBJECTS of inquiry.

Q: How do you describe the abstract algebra course you have taken recently?

S1: Well, ... lot of [it] is just weird things. You end up learning things like $1+1$ is not 2 anymore....

Q: Explain the argument made by the statement: $(-1)x = -x$

S2: There is no argument. It is obvious that negative one times x is negative x.

Fig. 2.

Table 2

	N	%
Contextual (General vector space statements are interpreted and proved in the contest of $\mathbb{R}^n$)	33	66
Structural	10	20
Undecided	7	14
Total	50	100

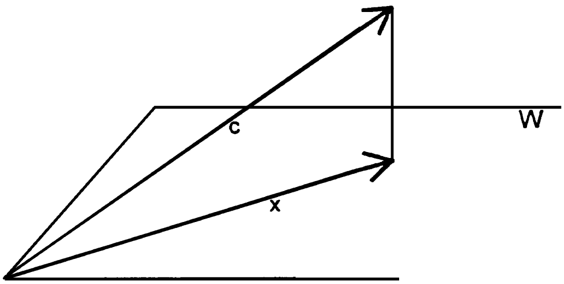

Fig. 3.

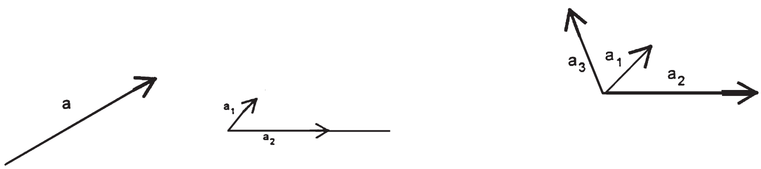

Fig. 4.

2. Category 2: students' understanding of mathematics in relation to the role of aristotelian causality in the mathematics of the renaissance

We tend to associate misconception and missing conceptions only with mathematically weak students. But in fact, all students, the weak and the able, in their desire to understand and make sense of the mathematical concepts we intend to teach them, encounter difficulty, and demonstrate as a result behaviors that in many cases are difficult to explain. Figs. 6 and 7 present examples of proofs to which certain students – always the more able students in the class – respond in a manner that has perplexed me. When the proofs are repeated to these students, they seem to understand each step in the proof, but at the end they reiterate the same question. What is the conceptual base for

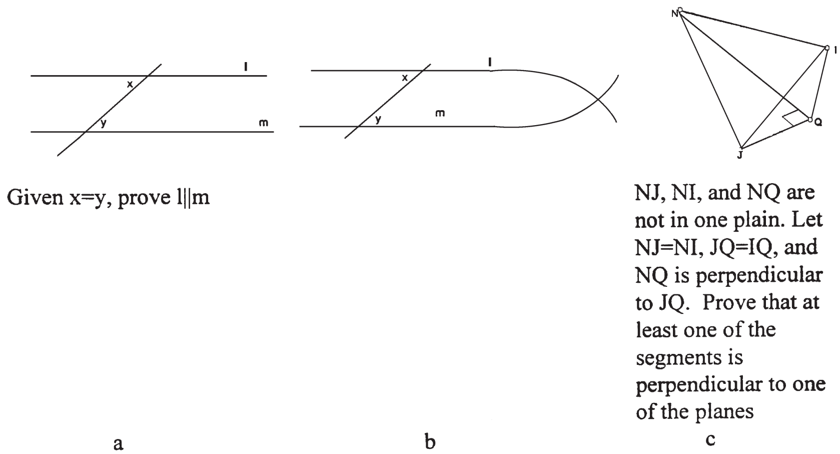

Fig. 5.

these responses? What really is the question these students are asking? While further research is needed to answer these questions, in what follows I will offer a conjecture.

Before I suggest what might be hidden in these to-us-strange responses, let me turn again to history:

"We do not think we understand something until we have grasped the why of it. ... To grasp the why of a thing is to grasp its primary cause", asserts Aristotle in *Posterior Analytics.* Mathematics is not a perfect science, argued 16–17th century philosophers, because an "implication" is not just a logical consequence; it must also demonstrate the *cause* of the conclusion. Some

Problem: Prove that any three vectors in R^2 are linearly dependent.

Proof (given by the instructor)

Let $\begin{bmatrix} a \\ b \end{bmatrix}, \begin{bmatrix} c \\ d \end{bmatrix}, \begin{bmatrix} e \\ f \end{bmatrix}$ be three vectors in R^2, and consider the system $AX = 0$, where

$A = \begin{bmatrix} a & c & e \\ b & d & f \end{bmatrix}$. The system $AX = 0$ has at least one free variable; therefore, it has a non-zero solution, $X = [x_1 \quad x_2 \quad x_3]$. Since $x_1A^{(1)} + x_2A^{(2)} + x_3A^{(3)} = 0$, and x_1, x_2, x_3 are not all zero, one of the columns of A must be a linear combination of the others. Hence the columns of A are linearly dependent.

Some students respond by saying something to the effect what if the system weren't homogeneous? Your answer is dependent on the fact that the system is homogeneous. If the system weren't homogeneous, you wouldn't be able to prove that the vectors are dependent.

Fig. 6.

Theorem. Let T be a square matrix, and let $v_1, v_2, \ldots, v_k$ be eigenvectors of T that correspond to eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are all distinct, then $v_1, v_2, \ldots, v_k$ are linearly independent.

Proof:

Assume $\sum_{i=1}^{k} a_i v_i = 0$. We should show that $a_i = 0$ for each i. Let $p_i(x) = \prod_{j \neq i}^{k} \frac{(x - \lambda_j)}{(t_i - t_j)}$ for $i = 1, 2, \ldots, k$. By the Spectral Mapping Theorem,

$$0 = p_i(T)0 = p_i(T)\sum_{j=1}^{k} a_j v_j = \sum_{j=1}^{k} p_i(T) a_j v_j = \sum_{j=1}^{k} a_j p_i(T) v_j = \sum_{j=1}^{k} a_j p_i(\lambda_j) v_j$$

Since $p_i(\lambda_j) = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$, $0 = p_i(T)0 = \sum_{j=1}^{k} a_i \delta_{ij} v_i = a_i v_i$. Hence, $a_i = 0$ for each i.

Some students respond by saying something to the effect what if you chose different polynomials? $v_1, v_2, \ldots, v_k$ are independent because you chose these polynomials. Maybe if you have chosen polynomial other than these, you wouldn't be able to prove the theorem.

Fig. 7.

mathematicians (e.g., Barozzi, 16–17th century) argued that some parts of mathematics are more scientific (causal) than others; but that proof by contradiction is not a causal proof, and therefore it should be eliminated from mathematics. Others (e.g., Barrow, 16–17th century) argued that all mathematics proofs are causal including proof by contradiction (see [2]).

To illustrate the nature of this debate, consider the Euclid's Proposition 1.32 and its proof:

The sum of the three interior angles of a triangle is equal to 180°.

Proof. Construct CE parallel to AB (Fig. 8). Then the alternate angles BAC and ACE are congruent and the corresponding angles ABC and ECD are congruent. Hence,

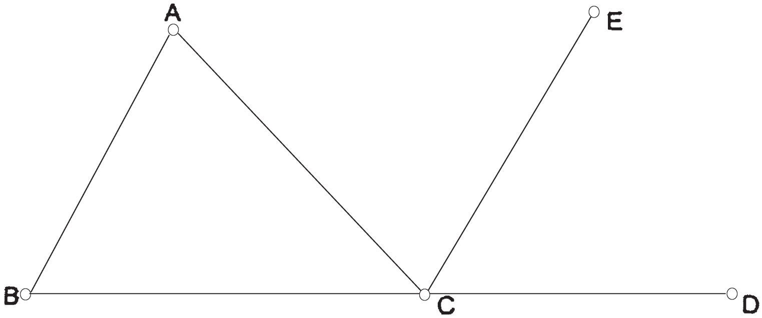

Fig. 8.

$$m(\mathrm{ABC}) + m(\mathrm{BAC}) + m(\mathrm{ACB}) = m(\mathrm{ECD}) + m(\mathrm{ACE}) + m(\mathrm{ACB}) = 180.$$

What is the cause of the property that is proved here, asked these philosophers? The proof appeals to two facts about the auxiliary segment CE and the external angle ACD. But these facts, they argued, cannot be the true cause of the property. For the property holds whether or not the segment CE is produced and the angle ACD considered.

The scientificaness of mathematics was denied primarily based on the pervasive use of proofs by contradiction, which, in the eyes of many philosophers, did not qualify as causal proofs. When a statement "A implies B" is proved by showing how NOT B (and A) leads logically to an absurdity, we do not learn anything about the causality relationship between A and B. Nor, continued these philosophers to argue – do we gain any insight of how the result was obtained [2].

Another interesting argument against the scientificaness of mathematics was this: If mathematical proof were scientific (i.e., causal), then proofs for "A if and only if B" statements entail that A is the cause of B and B is the cause of A. Hence, A is the cause of itself, which is absurdity, because nothing can be its own cause.

Since the basis of proof by exhaustion is proof by contradiction, it too was unsatisfactory to many mathematicians of the 16th and 17th centuries. They felt that the ancients, who broadly used proof by exhaustion to avoid explicit use of infinity, failed to convey their methods of discovery.

Were these issues of marginal concern to the mathematics of the sixteen and seventeen centuries, or had they significantly affected it? To what extent did the practice of mathematics in the 16 and 17 centuries reflects a global epistemological positions that can be traced back to Aristotle's specifications for perfect science? These are important questions, if we are to draw a parallel between the individual's epistemology of mathematics and that of the community.

Mancosu [2] argues that the practice of Cavalieri, Guldin, Descartes, and Wallis, and other important mathematicians reflects a deep concern with these issues. He shows, for example, how two of the major works of the 1600s – the work by Cavalieri on indivisibles and that by Guldin, his rival, on centers of gravity – aimed at developing mathematics by means of direct proofs. These two mathematicians, argued Mancosu, explicitly avoided proofs by contradiction in order to conform to the Aristotelian position on what constitutes perfect science – a position Aristotle articulated in his *Posterior Analytics*.

Descartes, whose work represents the most important event in 17th-century mathematics, was heavily influenced by these developments. Descartes appealed to a priori proofs against proofs by contradiction because they show how the result is obtained and why it holds, and they are *causal* and ostensive.

Before we go back to the students' responses in Figs. 6 and 7, let me present one more observation. A group of eight inservice teachers were presented with two proofs of Proposition 1.32: the proof just presented and the following proof which was originally offered by a preservice teacher (Amy) taking a course in college geometry (see [1]):

> Amy demonstrated to the whole class how she imagines the theorem, "The sum of the measures of the interior angles in a triangle is 180°." Amy said something to the effect that she imagines the two sides AB and AC of a triangle ABC being rotated in opposite directions through the vertices B and C, respectively, until their angles with the segment BC are 90° (Fig. 9a and b). This action transforms the triangle ABC into the figure A′BCA, where A′B and A″C are perpendicular to the segment BC. To recreate the original triangle, the segments AB and A″C are tilted toward each other until the points A′ and A″ merge back into the point A (Fig. 9c). Amy indicated that in doing so she "lost two pieces" from the 90° angles B and C (i.e., angles A′BA and A″ CA) but at the same time "gained these pieces back" in creating the angle A. This can be better seen if we draw AO perpendicular to BC: angles A′BA and A″ AC are congruent to angles BAO and OAC, respectively (Fig. 9d).

All eight teachers preferred Amy's proof to the standard Euclid's proof, saying that it shows *why* the sum of the angles in a triangle is 180°. They indicated that through Amy's proof they could see how the construction of the triangle "made" the sum of the angles 180°. For these teachers, I suggest, Amy's proof was a *causal* proof – an enlightening proof that gives not just mere evidence for the truth of the theorem but the *cause* of the theorem's assertion.

The history of the development of the concept of proof, as I have briefly reviewed here, suggests that our current understanding of proof was born out of an intellectual struggle during the Renaissance about the nature of proof – a struggle in which Aristotelian causality played a significant role. If we assume that the epistemology of the individual mirrors that of the community, we

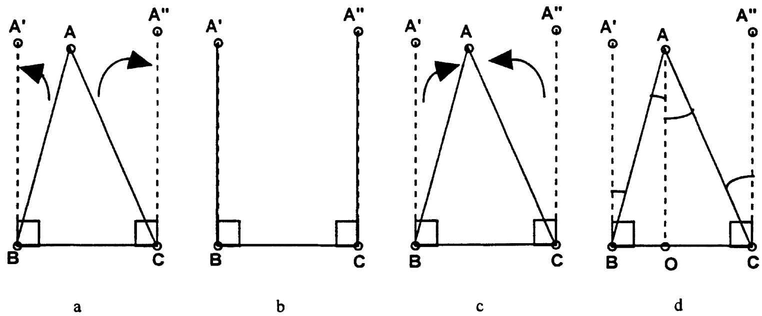

Fig. 9.

should expect the development of students' conception of proof to include at least some of the major obstacles encountered by the mathematics community through history. I conjecture that Aristotelian causality is one of these obstacles. Causality is more likely to be observed with able students, who seek to understand phenomena in depth, than with weak students, who usually are satisfied with what ever the teacher presents. It is possible, for example, that the students' responses in Figs. 6 and 7 are a manifestation of the causality phenomenon. The students who responded to the proof in Fig. 6 by saying "What if the system weren't homogeneous?" had interpreted the homogeneous system $AX = 0$ to be the *cause* for the independency of the vectors

$$\begin{bmatrix} a \\ b \end{bmatrix}, \begin{bmatrix} c \\ d \end{bmatrix}, \begin{bmatrix} e \\ f \end{bmatrix}$$

and so they desired to understand the exact causality relationship. Similarly, the students who responded to the proof in Fig. 7 by saying "What if you took different polynomials" sought to understand the cause-effect relationship between the Lagrange polynomials,

$$p_i(x) = \prod_{j \neq i}^{k} \frac{(x - \lambda_j)}{(t_i - t_j)}$$

and the theorem's assertion about the independency of the eigenvectors.

3. Implications

The observation regarding causality is still in its infancy. Systematic studies on the existence of this phenomenon with students must be conducted before any conclusion can be drawn.

On the other hand, our data strongly support the observation regarding students' difficulties extending their proof understandings – from the contextual proof to the structural proof and axiomatizing proof – thus, my conclusions will be based on this observation alone.

1. I have doubts about the wisdom of starting off college geometry courses with finite geometries. Our data suggest that a more promising approach would be to begin with axiomatic Euclidean geometry, which corresponds to and extends students' physical reality.

2. For high school geometry, I believe it is imperative to reinstate synthetic Euclidean geometry in high school mathematics curricula, whose place has recently been significantly eroded. Synthetic Euclidean geometry in 2 and 3 dimensions can serve as an intermediate stage to an advanced level of mathematical thinking, particularly, to the structural conception of proof.

An obvious justification for this assertion is that Synthetic Euclidean geometry is a concrete context in which students can learn the concept of a deductive system – a concept college students largely lack. I will not discuss this justification here but focus instead on a less obvious, but critically important justification.

We have seen earlier how students are unable to detach from their spatial imageries in dealing with abstract algebraic structure – a phenomenon we called *contextual* conception. One manifestation of this conception is that students do not understand the role of symbolic figures (geometric illustrations) in algebraic structures such as $\mathbb{R}^n$. Synthetic Euclidean geometry is an excellent context to teach students to view symbolic figures as *representations* of "reality", not the "reality" itself. For example, in proving the statement in Fig. 5a by proof-by-contradiction, one must first assume that lines l and m intersect (Fig. 5b). This usually causes a discomfort to many students because the figure conflicts with their image of a straight line, and they begin to realize that the figure is a *representation* of a hypothetical spatial situation. Similarly, Fig. 5c is a source of difficulties to many students because the properties the student is asked to prove are not directly seen in the figure but needs to be visualized. Again, here too the students begin to learn that the figure is just a representation of reality, not the reality itself.

It is interesting to note that according to our data, the ability to deal with 3-dimensional Euclidean geometry correlates with the ability to think axiomatically (as in finite geometries).

3. My conclusions for the teaching of linear algebra might be controversial. Instructors and textbook writers have good intention when they start with geometric motivations in introducing new concepts. We all intend to use geometry to provide students with geometric insight. But, I am afraid, these geometric illustrations do not achieve their purpose. On the contrary, they hinder students learning the true concepts of linear algebra.

The sequence in which we present material to students and the way we introduce new concepts are critical learning factors. When geometry is introduced before the concept has been formed, the students view the geometry as the raw material to be studied, they remain, as a result, in the restricted world of geometric vectors, and do not move up to the general case.

As we should be careful not to move students up hastily from $\mathbb{R}^n$ to more general vector spaces, we should be as careful how to introduce special cases of $\mathbb{R}^n$, namely the geometry of directed segments. The student must stand on solid ground as to the world he or she is studying. In elementary linear algebra, there should be one world: $\mathbb{R}^n$ – at least during the early period of the course.

Having said this, I want to emphasize and reemphasize that I am not advocating the elimination of geometry from linear algebra. On the contrary, geometry can be a very powerful tool in solidifying concepts the students have formed or begun to form. I am only raising questions about its usefulness in motivating new linear algebra concepts.

There are at least three reasons that make geometric motivations very popular among textbooks and instructors of linear algebra.

The first reason is that the instructor sees how the geometric situation is isomorphic to the algebraic one and so he or she believes that the geometric concept can be a corridor to the more abstract algebraic concept. The problem with this preconception is that the student does not share this important insight.

The second reason is that the geometric concepts are relatively easy to understand, and so it is only natural to use them as startup ideas. For example, the concept of linear independence is commonly introduced via collinear directed segments. The problem with examples such as this is that they are very simple and easily understood by every student, whereby the students form an extremely powerful concept image that it is hard for many to relinquish.

The third reason is that the instructor well understands that the geometric illustrations are only introductory ideas to the abstract concepts that are yet to come. He or she knows that the true labor is to abstract the concepts from the geometric ideas and has developed the patience to do so. Unfortunately, a regular student does not have this basic metamathematical understanding.

References

[1] G. Harel, L. Sowder, Students' proof schemes, in: E. Dubinsky, A. Schoenfeld, J. Kaput (Eds.), Research on Collegiate Mathematics Education, vol. III, American Mathematical Society, 1998.

[2] P. Mancosu, Philosophy of Mathematical Practice in the 17th Century, New York, Oxford University Press, 1996.

[3] J. Klein, Greek Mathematical Thought and The Origin of Algebra, MIT Press, Cambridge, MA, 1968.