* **CHAPTER 12** *

# Expert Thinking and Structure Theorems

In this chapter we return to the journey through the axiomatic formal world. In chapter 8 we saw the complication involved in the initial stages of formal deduction of a relationship such as $(-a)(-b)=ab$ in an appropriate axiomatic system. In chapter 10 we saw that first stage of dealing with definitions and deductions is highly complicated as learners attempt to make sense of the formal ideas when their minds are already full of embodied and symbolic ideas that now must be reorganized into formal definitions and proof.

Learners may develop in a variety of ways—as natural learners building structurally on embodied mental images of situations, or operationally on experiences manipulating symbols, or in a more formal way based on making deductions from formal definitions. Some may learn proofs procedurally to reproduce them in examinations.

When a learner is presented with a list of axioms, the first stage is to prove some initial theorems that enable the axioms and definitions to be used in more flexible ways. This leads from a multistructural list to a growing relational structure of formal knowledge. In the longer term this leads to the emergence of crystalline concepts.

For instance, a complete ordered field is formulated with a list of axioms including the axiom of completeness in the context of an ordered field. In chapter 10 the completeness axiom was given in the form 'a cauchy sequence always tends to a limit'. There are other ways of formulating this same axiom for an ordered field, for example, 'an increasing sequence bounded above converges to a limit', 'a decreasing sequence bounded below converges', 'a non-empty set bounded above has a least upper bound', or 'a non-empty set bounded below has a greatest lower bound'. All of these different versions can be proved to be equivalent in the context of the axioms for an ordered field.

The human mind can compress the ideas further, not only into a concept in which different versions are 'equivalent', but also into a single crystalline structure: the one and only complete ordered field $\mathbb{R}$ that has formal, embodied and symbolic aspects all blended into one.

In the case of other formally defined mathematical concepts, such as a group, the concept may exist in many different forms. Examples include the group of rational numbers under addition, the non-zero rationals under multiplication, a group of transformations, or in a multitude of other examples. In group theory we have the notion of isomorphism to describe groups with a bijection (one-one onto map) between the sets that respects the group operation. Isomorphic groups are fundamentally the

same and the human brain can conceive such a concept as a single crystalline structure.

Even the general concept of group has its own crystalline structure that students begin to grasp. This is illustrated by a conversation that I had with a group of thirty mathematics students beginning a course on 'the development of mathematical concepts' after a year of formal university mathematics.

Few of the students claimed to have fully understood the first year group theory course. Some claimed that they had learnt virtually nothing. I took a positive attitude by inviting the sceptics to respond to a challenge. I told them that I had a group $G$ in my mind. I wouldn't tell them any more than that, but I wanted them to tell me about it.

The first suggestion was that it must have an identity element that we agreed to call $e$. When asked about other elements, after a suggestion that perhaps $e$ was the only element, I asked what could be said about any other element, say $x$. The response came that we could multiply $x$ by $x$ and get $x^2$. I then enquired about multiplying $x^2$ by $x^2$ and was given the immediate answer $x^4$. The explanation was given that this used the associative law so that, either way, the result is $x$ times $x$ times $x$ times $x$.

The point had now been made. While the students did not know the group I had in mind, they had a sense of its properties and could talk to me about it, and this information followed from the growing knowledge structure each of them had constructed for the group concept.

In this chapter we will consider the longer-term development in which multistructual lists of axioms lead to theorems that relate properties and compress knowledge into crystalline formal concepts.

The journey includes the proof of a special kind of theorem, called a *structure theorem*, which proves consequences that reveal new embodied forms with precise formal meanings and symbolic methods of operation that are an inevitable consequence of the axioms. This transforms the whole nature of formal mathematical thinking. Instead of an initial multistructural system of axioms from which theorems must be proved carefully step-by-step, the developing expertise leads to crystalline concepts that have new forms of embodiment and symbolism now underpinned by the power of formal proof.

## COMPARING NOVICE AND EXPERT

Keith Weber compared the ways in which four undergraduates and four doctoral students solved formal mathematical problems in group theory.[1] He found that the undergraduates were able to reproduce simple theorems but were unsuccessful when the problems became more complex. Meanwhile, 'the doctoral students appeared to know the powerful proof techniques in abstract algebra, which theorems are more

important, where particular facts and theorems are likely to be useful, and when one should or should not try and prove theorems using symbolic manipulation.' [2]

A typical question was to prove whether or not the group of integers $\mathbb{Z}$ under addition was isomorphic to the group $\mathbb{Q}$ of rational numbers addition. In terms of a definition, this would mean finding a bijection between $\mathbb{Z}$ and $\mathbb{Q}$ that preserved addition, or showing that no such bijection existed. None of the four undergraduates could provide a formal response but all four graduates were able to do so.

The undergraduates focused on their memory that $\mathbb{Z}$ and $\mathbb{Q}$ have the same cardinal number and so already had a bijective correspondence between them. They had part of the idea—a bijection—but not a bijection that respected the operation of addition. The powerful met-before of the equal cardinality of $\mathbb{Z}$ and $\mathbb{Q}$ was so strong that they were unable to make any further progress.

The graduates had richer knowledge structures supporting their thinking. For instance, one immediately declared that $\mathbb{Q}$ and $\mathbb{Z}$ could not be isomorphic, first by speculating that $\mathbb{Q}$ is dense but $\mathbb{Z}$ is not, then that '$\mathbb{Z}$ is cyclic, but $\mathbb{Q}$ is not' (meaning that the element 1 generates the whole of $\mathbb{Z}$ under addition but no element in $\mathbb{Q}$ does so). This second insight shows that the additive groups of $\mathbb{Z}$ and $\mathbb{Q}$ cannot be isomorphic.

The undergraduates still saw proof as a process and sought a process to establish a bijection or not. The graduates had developed a richly connected formal knowledge structure and knew that isomorphic groups would have the same properties and used this to solve the problem.

## THE PROCESS OF PROVING AND WARRANTS FOR TRUTH

In the process of shifting from the initial workings with definitions to the building of a succession of theorems to give a rich knowledge structure, the prover must have some idea as to what theorems might be worth proving and to conjecture what might be true. This involves producing some kind of argument to support the proposed theorem, which may not initially be a formal proof but, nevertheless, gives the prover increasing conviction that the theorem is true.

In the 1950s, the philosopher Toulmin[3] considered how general arguments were composed. In general, a proof might consist of some given **Data** that was assumed true and a form of **Proof** to establish the **Conclusion**. (Figure 12.1.)

*Figure 12.1: Proving a conclusion from given data*

Toulmin, however, saw that arguments were more general than this. He suggested that a general argument began with **Data**, then had a **Warrant** for truth in the forms of arguments that support the likelihood of truth, without necessarily giving a 100% certainty, and as a consequence there would be a **Qualifier** expressing the degree of confidence in the argument supporting the **Conclusion**. The Warrant was supported by **Backing**, in the form of additional evidence, but the argument may have a **Rebuttal** which potentially refutes the conclusion, for instance, by stating conditions in which the argument would fail. (Figure 12.2.)

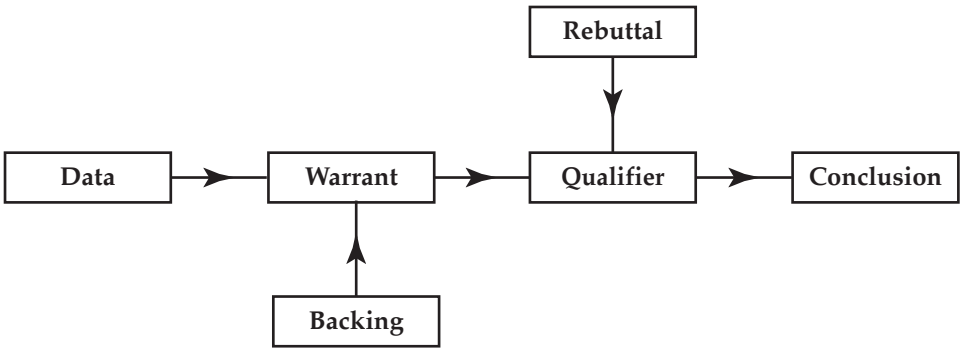


*Figure 12.2: Toulmin's model of general argument*[4]

This framework gives a wider context for describing how new proofs are suggested by individuals using their current knowledge structures.

Matthew Inglis and his colleagues[5] gave high quality mathematics research students unfamiliar problems in clinical interviews to study how they developed their arguments. These concerned variations of the definition of a perfect number, given as follows for a positive integer $n$:

> A *perfect number n* has divisors (including 1 and $n$) that add up to $2n$,
> An *abundant number n* has divisors that add up to more than $2n$,
> A *deficient number n* has divisors that add up to less than $2n$.

For instance, 6 is perfect (because $1+2+3+6=12$), 7 is deficient (because $1+7=8<14$) and 12 is abundant (because $1+2+3+4+6+12=28>24$).

Amongst the conjectures that were presented to the students to test if they were true or false were the following:

(A) The sum $m+n$ of two abundant numbers $m$, $n$ is abundant,

(B) The product $mn$ of two abundant numbers $m$, $n$ is abundant.

Graduate student Chris responded immediately to (A) saying

> That doesn't look true. … Because the factors of $m+n$ don't really have anything to do with the factors of $m$ or $n$. So it should be fairly easy to construct a counterexample.

He went on to find two abundant numbers whose sum was not abundant.

Looking at (B), however, he said

> Right, so if $m$ and $n$ are abundant then $mn$ is abundant. That looks more plausible, because they are going to share factors.

This led to a sequence of arguments that ended up with a proof. The reader might like to solve both problems to see if they are, in fact, true or false. This will involve the kind of thinking that is being discussed in this section and is likely to give deeper insight than just passively reading about the ideas.

The important observation to make is that initially the responses give warrants that are qualified, in the first case a sense that it should be 'fairly easy' to prove false, the second looks 'more plausible' with the backing that the numbers are going to share factors.

A study of the responses of the six students suggested three different types of warrant:

> An *inductive* warrant-type based on evaluating one or more specific cases.
>
> A *structural-intuitive* warrant-type based on observations about or experiments with some kind of mental structure, be it visual or otherwise, that persuades them of a conclusion.
>
> A *deductive* warrant-type using formal mathematical justifications to warrant the conclusion, including deductions from axioms, algebraic manipulations, or the use of counterexamples.

These three types correspond to

> the use of examples, typical in the first stage of problem solving (and also in the earlier pre-axiomatic stages of proof development),
>
> the use of an imagined mental structure that may be 'visual or otherwise',
>
> an argument that is essentially formal deductive from axioms or algebraic manipulation using quantified statements in generalized arithmetic.

In the three-world framework, the first and second represent stages of development in embodiment and symbolism, with embodied images or calculations, while the third sees a switch to formal proof, or an operational manipulative proof using quantifiers. The given problem is based on arithmetic operations with numbers that could be tackled without a formal knowledge framework. Here it is tackled by talented graduate students experienced in formal proof using their sophisticated knowledge structures to provide initial warrants for truth before moving on to seek a formal proof.

Pablo Meija-Ramos investigated a range of other problems that carried different levels of conviction from specific examples, embodied pictures, symbolic manipulations and formal proofs.[6]

Two conjectures given to students were the following:

(C) The derivative of an even function is an odd function,

(D) The product of two diagonal matrices is diagonal.

The first can be imagined as an embodied picture, a symbolic relationship between functions, or something related to the formal limit definition of the derivative. The second focuses more on the symbolism of matrix multiplication. In practice, the first produced a range of responses, for example, using the symmetry of the even function $f$ satisfying $f(x) = f(-x)$ to embody the slopes as mirror images, with the slope $f'(x)$ equaling minus $f'(-x)$. This insight was usually considered to be meaningful but did not carry the full conviction of proof. (Figure 12.3.)

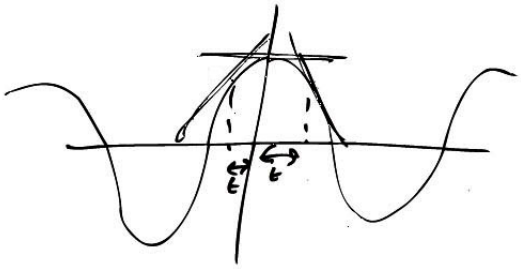


*Figure 12.3: dynamic visual picture of slope of an even function at $x = -t$ and $x = t$*

The argument produced by one such student was analyzed using the Toulmin diagram as in Figure 12.4.

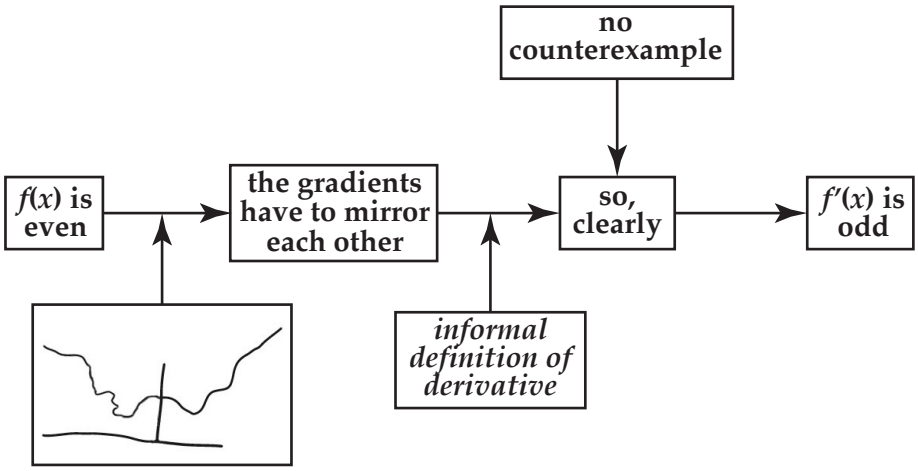


*Figure 12.4: The Toulmin analysis of a visual proof by one of the students*

Here the student uses the picture as a backing to produce the warrant that the gradients have to mirror each other, with an implicit use of an informal definition and no counterexample as a rebuttal to suggest that the conclusion 'clearly' follows. When questioned, the student confirmed that although the argument was 'clear', it was not a formal proof.

Overall, the students gave a variety of arguments that were classified as 'inductive arguments' (for instance looking at special cases such as specific polynomials with even powers), 'informal deductive arguments' (such as that in figure 12.3) or 'formal deductive arguments'. These arguments are respectively operational symbolic, structural embodied thought experiment, and formal proof.

The case of the diagonal matrix is different as it is expressed symbolically without an obvious embodied representation. Yet even here, there are some solutions that recall the definition of matrix multiplication as a *functional* embodiment, where the rule of multiplication combines a row in the first matrix and a column in the second by multiplying the successive entries (along the row and down the column), adding them together and placing the result in the position where the row and column cross each other. This calculation was accompanied by a drawing (figure 12.5) in which enactive gestures traced the directions along row and column to give an iconic representation of the rule for multiplication. It enabled the student to build up the formula for the product of two matrices to shift from an informal deductive argument to a formal deductive argument.

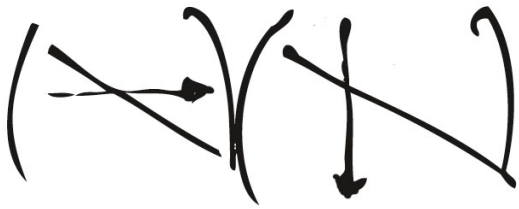

*Figure 12.5: matrix multiplication in outline*

Mejia-Ramos summarized his data in the following terms:

> Students may use any type of argument available to them when approaching a given task, and that while this may result in certain students using the same combination of arguments in two different tasks (empirical to estimate the truth of the conjecture, informal deductive to explain it and formal deductive to prove it), others may use only a formal deductive approach when such an argument is available to them, but resort to other types of arguments when faced with tasks for which the basis of a symbolic argument is not as 'fresh in their minds.'[7]

Both of these studies are designed to use the Toulmin framework to reveal the need for a qualifier that expresses a certain level of confidence rather than an all-or-nothing proof. As such they focus on the development of a formal argument, rather than the specific nature of embodiment or symbolism involved. However, the roles of embodiment and symbolism are clear in the examples given, which each example having its own specific characteristics. (A) and (B) are general properties of whole number arithmetic. (C) is a calculus problem that can be

embodied as a visual picture, symbolized as a rule in calculus, or formalized in mathematical analysis. (D) is a problem in matrix algebra that is essentially symbolic. Hence (A) and (B) are problems requiring a symbolic proof in whole number arithmetic, (C) offers distinct arguments in embodiment, symbolism or the formalism of mathematical analysis, and (D) requires a proof in the formal symbolism of matrix algebra, but involves aspects of functional embodiment to remember the rule for matrix multiplication. Each benefits from different forms of support in embodiment, symbolism and formalism to construct a formal proof.

## STRUCTURE THEOREMS AND NEW FORMS OF EMBODIMENT AND SYMBOLISM

The axioms and definitions proposed for formal concepts are chosen by the mathematician to express generative properties that can be used as a foundation for formal development. Such axiomatic structures rarely reflect *all* the properties that are in the original examples that inspired the theory. For instance, a vector in elementary mathematics arises from generalizing the notion of vector in two and three dimensions as a quantity with magnitude and direction. But a vector space is not axiomatized to suggest this structure at all. Instead the axioms of a vector space focus only on the properties of addition of vectors and operations on vectors by elements from the associated field.

In the following discussion, the formal definition of various mathematical systems will be analyzed to show that, under specified circumstances, a formally defined concept will have structural properties that endow it with a specific embodiment and corresponding symbolism.

The ideas will be outlined in a general fashion. What is important for the general reader is to grasp that the formal definitions leads to embodied and symbolic properties. The mathematician may choose the axioms to suit specific purposes, but those axioms then have inevitable properties that result in a specific crystalline structure.

## THE CONCEPT OF VECTOR

A vector space $V$ over a field $F$ is defined formally as a commutative group $V$ with the operation combining $\mathbf{u},\mathbf{v}\in V$ written as $\mathbf{u}+\mathbf{v}\in V$ and multiplication of a vector $\mathbf{v}\in V$ by $a\in F$ denoted as $a\mathbf{v}\in V$ that satisfies the axioms: $(a+b)\mathbf{v}=a\mathbf{v}+b\mathbf{v}$, $a(\mathbf{u}+\mathbf{v})=a\mathbf{u}+a\mathbf{v}$, $(ab)\mathbf{v}=a(b\mathbf{v})$ and $1\mathbf{v}=\mathbf{v}$.

An example of a vector space is the $n$-dimensional space $F^n$ of coordinate vectors $(a_1,\ldots,a_n)$ where the coordinates $a_1$, …, $a_n$ are all elements in the field $F$. This is a generalization of two-dimensional real space $\mathbb{R}^2$ or three-dimensional real space $\mathbb{R}^3$ that can be embodied as vectors in the real plane or in three-dimensional real space. Addition of

vectors simply adds coordinates and multiplying by an element of $F$ multiplies each of the coordinates by that element.

To link the formal embodiment more closely to this example, further definitions are introduced.

A vector space is said to be *finite dimensional* if it has a finite set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ (called a 'spanning set') so that every vector $\mathbf{v}$ can be written as a combination $\mathbf{v} = a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n$ (for $a_1, \ldots, a_n \in F$). A set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is said to be *linearly independent*, if the sum $\mathbf{v} = a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n$ can only equal zero when $a_1 = \ldots = a_n = 0$.

If a spanning set is also linearly independent, then the representation $\mathbf{v} = a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n$ is *unique* in the sense that if there were two different expressions, $\mathbf{v} = a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n = b_1\mathbf{v}_1 + \ldots + b_n\mathbf{v}_n$, then

$$(a_1 - b_1)\mathbf{v}_1 + \ldots + (a_n - b_n)\mathbf{v}_n = 0$$

so $a_1 = b_1, \ldots, a_n = b_n$.

A set of vectors that spans a vector space and is also linearly independent is called a *basis*. In this case it can be proved that any two bases of a given vector space $V$ have the same number of elements and this is defined to be the *dimension* of the vector space.

In a vector space of dimension $n$, choose a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$, then any vector $\mathbf{v}$ can then be written uniquely as $\mathbf{v} = a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n$ which corresponds to the row vector $(a_1, \ldots, a_n)$ in the vector space $F^n$.

This leads to the essential structure theorem that any finite-dimensional vector space over a field $F$ is isomorphic to $F^n$, or, to put more simply, that an $n$-dimensional vector space can have its elements represented as $n$-tuples $(a_1, \ldots, a_n)$ where the coordinates $a_1, \ldots, a_n$ are all elements in the field $F$. In particular, if $F$ is the field $\mathbb{R}$ of real numbers, then a two or three-dimensional vector space over $\mathbb{R}$ is isomorphic to two or three-dimensional cartesian space.

This structure theorem not only asserts that any finite-dimensional vector space over a field $F$ is isomorphic to $F^n$, it also opens the doors to link the formal structure to the worlds of embodiment and symbolism. Vectors in finite-dimensional spaces can be written using coordinates (usually as column vectors), linear maps can be written as matrices and multiplied using the symbolism of matrix multiplication. It reveals the underlying crystalline structure for a finite-dimensional vector space that is already familiar from matrix algebra.

If the field $F$ is the field of real numbers $\mathbb{R}$, then a three-dimensional vector space $V$ has the structure of vectors in $\mathbb{R}^3$, except that there is as yet no definition of angle between vectors (which requires additional axioms). Even so, it enables vector subspaces to be seen as lines or planes through the origin. The structure theorem now blends the formal axiomatic structure with its familiar embodiment in space and its linear functions as symbolic matrices. It yields an inevitable structure blending

embodied, symbolic and formal aspects into a single crystalline concept.

In chapter ten, it was suggested that students construct formal mathematical knowledge either by a *natural* process of *giving meaning* to the formal definition using embodiment and symbolism, or by a *formal* process of *extracting meaning* from the definition by reflecting on how it operates in formal proofs.

A structure theorem, as proved here for finite dimensional vector spaces, extends the possibilities of formal proof by giving formal structures new kinds of embodiment and symbolism that can be used to imagine new possible theorems and their warrants for truth that can become part of the proving process (extending the previous figure 10.11 to figure 12.6).

In mathematical research, mathematicians have highly sophisticated knowledge structures built from experience in producing proofs. However, inventing and proving new theorems remains a problem-solving process, which may build naturally on expert knowledge structures that include blends of embodiment, symbolism and formalism to suggest conjectures that may later be proved formally.

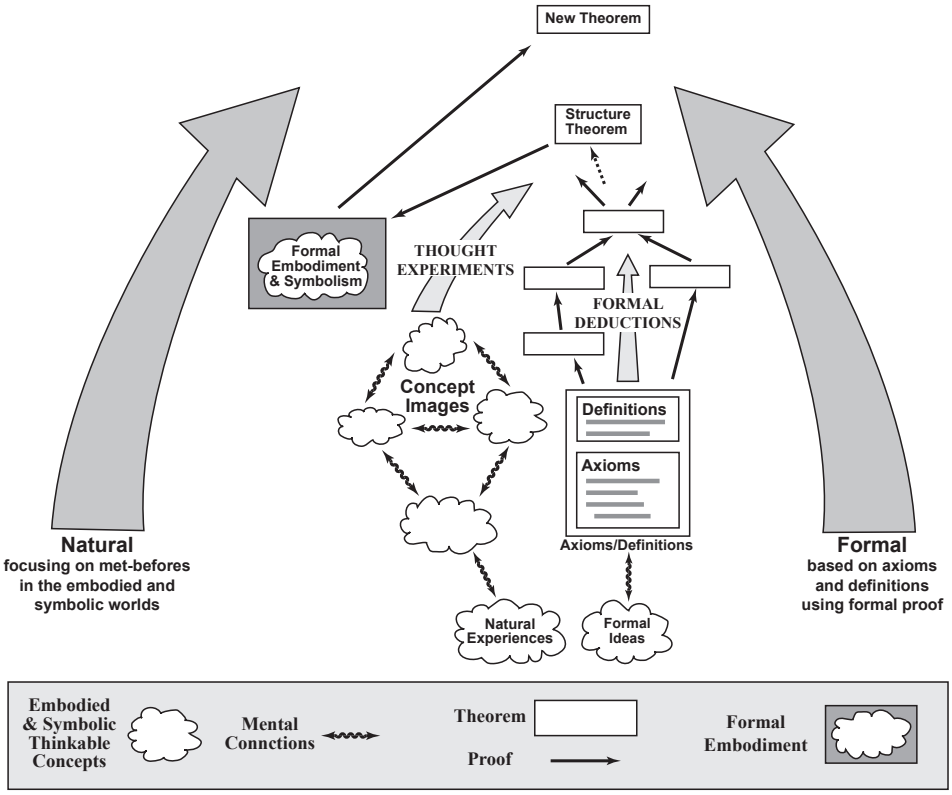


*Figure 12.6: Structure theorems enhance formal proof with embodiment and symbolism*

In this framework, structure theorems play a significant role by underpinning formal theory with supportive embodiment and symbolism. They provide a natural structure for contemplating new possibilities suggested by the crystalline structure of the formal concepts and a working environment for the application of ideas.

They also provide practical ways in which formal mathematics may be used in applications. Applied mathematicians use the embodiment and symbolism to formulate problems by imagining the embodiment and formulating a symbolic model to solve the problem, without consciously being concerned with the formalism that is now firmly established.

The following sections outline more structure theorems in other areas of mathematics. They are seminal in giving a formal underpinning to the human blending of embodiment, symbolism and formalism.

## A FINITE GROUP IS ISOMORPHIC TO A GROUP OF PERMUTATIONS

Practical examples of groups include groups of permutations, where the elements permute (or re-order) the elements in a given finite set. The group of permutations of $n$ elements is called the *symmetric group of order* $n$ and denoted by $S_n$. It consists of all the functions permuting the numbers 1, 2, …, $n$. This can be used for computational purposes. For example, the permutation of the elements {1, 2, 3, 4, 5} that takes the order 12345 into 21453 takes 1 to 2, 2 to 1, 3 to 4, 4 to 5 and 5 to 3. This permutes the elements in two cycles, one taking 1 to 2 and 2 to 1, the other cycling 3 to 4 to 5 and taking 5 back to 3. (Figure 12.7.)

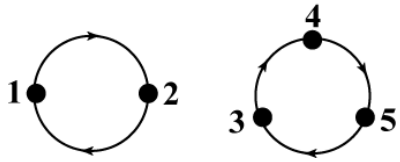


*Figure 12.7: A permutation represented as cycles*

These cycles are written sequentially in brackets as (12), (345) where the elements in a bracket are permuted, with each one going to the next except the last which goes to the first. Permutations are combined by performing one followed by the other, reading from left to right. Thus the product (123)(123) takes 1 to 2 in the first bracket, then 2 to 3 in the second, taking 1 to 3 overall, while 2 goes to 3 goes to 1 overall, and 3 goes to 1 goes to 2, giving the cycle (132). This permutation turns the cycle on the right of the figure two turns clockwise, and has the same effect as one turn anticlockwise. The product (123)(123)(123) turns through three turns clockwise, turning full circle to give the identity of the permutation group. This enables the operations in the symmetric group to be embodied and calculated using cyclic permutations.

Given any finite group $G$ with $n$ elements, then each element $g \in G$ gives a function $f_g : G \to G$ by multiplying every element on the left by $g$, to get $f_g(x) = g \circ x$. By the group axioms it can be shown that $f$ is a bijection that permutes the elements of $G$. Numbering the elements $g_1$, …, $g_n$, gives a permutation of the $n$ elements 1, 2, …, $n$, which is an element of the group $S_n$ of permutations. This assignment maps elements of $G$ to elements of $S_n$ and preserves the group operation because $f_h(f_g(x)) = h \circ g \circ x = f_{h \circ g}(x)$. This shows that any finite group $G$ can be represented as a subgroup of a group of permutations. Once this structure theorem is established, it allows the formal theory of finite groups to be linked to symbolic calculations.

However, mathematicians shift to a new level where equivalent structures are reconceptualized as a single crystalline structure that can be expressed in different ways. Using this way of thinking, we see a finite group not just as being *isomorphic* to a subgroup of a permutation group, but think of it *as* a subgroup of a permutation group.

Now expert mathematicians are in a completely new ball game. Finite groups are not just systems given by the axioms of a group, they can now be seen structurally as subgroups of a permutation group. If we wish to classify all finite groups, then they are subgroups of permutation groups and the abstract problem of classification becomes a more concrete exploration of finding the subgroups of permutation groups.

In practice, the permutation group $S_n$ becomes very large as $n$ increases and can be very complicated, so various new techniques are developed. However, they are now based on *knowing* that a finite group is a meaningful concrete structure as a subgroup of a permutation group rather than just some kind of abstract idea.

Subsequent theorems analyze the possible structures that may occur, leading to new techniques to classify all finite groups (up to isomorphism). The problem is still open, but several giant steps have been made. One involves the notion of a 'simple' group, which acts as a kind of foundational group from which all other finite groups are formed. The finite simple groups have now been fully identified, taking us a step further in the overall development of group theory.[8]

## AN EQUIVALENCE RELATION ON A SET PARTIONS IT INTO EQUIVALENCE CLASSES

The notion of equivalence relation on a set $S$ is formulated entirely by three axioms specifying that the relation is reflexive, symmetric and transitive. It can be given an embodied meaning by looking at the elements that are equivalent to each other. Take any element $x \in S$ and consider the subset $S_x$ of all the elements equivalent to $x$. This is called the *equivalence class* containing $x$.

Every element of $S$ lies in an equivalence class (for instance, $x \in S$ belongs to the equivalence class $S_x$). If two equivalence classes have an element in common, say $z \in S_x$ and $z \in S_y$ then $z \sim x$ and $z \sim y$ and, using the reflexive and transitive axioms, we find $x \sim y$ and deduce that $S_x$ and $S_y$ are the same. This shows that the equivalence classes are either identical or non-intersecting, with each equivalence class being written in many ways as $S_x$ where $x$ is any element of the class. The equivalence classes can then be embodied as a partition of the set into distinct subsets (Figure 12.8.)

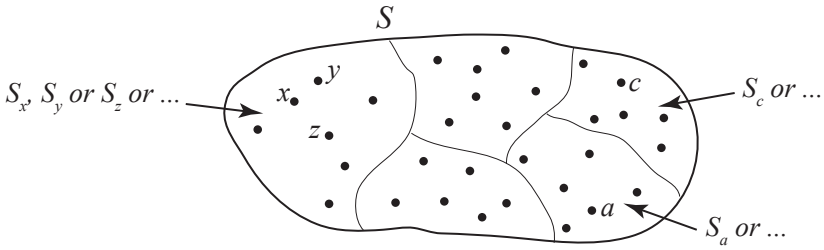


*Figure 12.8: Partitioning a set S using an equivalence relation*

Conversely, it is straightforward to show that any partition gives an equivalence relation. This gives a structure theorem in which an equivalence relation on a set, given in terms of the three axioms, reflexive, symmetric and transitive, can be embodied as partitioning the set into non-overlapping subsets.

This theorem reveals the wide generality of the notion of equivalence relation. Most of the equivalence relations met in algebra—equivalent fractions, equivalent algebraic expressions, equivalent vectors, equivalence in group theory, and so on—have the relation specified by a simple rule. But it need not be so. *Any* partition of a set gives an equivalence relation.

One equivalence relation that causes some problems to students is the idea of two infinite decimals being equivalent if they represent the same real number. Here the equivalence classes have one element in every case except where a finite decimal (such as 0.65) is equivalent to an infinite decimal ending in repeating nines (such as 0.64999…) where the equivalence class contains precisely two elements.

## AN ORDERED FIELD CONTAINS A SUBFIELD ISOMORPHIC TO THE RATIONAL NUMBERS

If a field $F$ is ordered, the order relation places a limitation on the arithmetic generated by the unit element $1_F$. Define the element $2_F = 1_F + 1_F$ and, proceeding by induction, for any $n \in \mathbb{N}$, define $(n+1)_F = n_F + 1_F$ to give the sequence $1_F$, $2_F$, …, $n_F$, … in $F$. In a general

field, the sequence might repeat, as happens in the field $\mathbb{Z}_3$ of integers modulo 3 where the sequence is $1_F$, $2_F$, $3_F$, where $3_F = 0_F$ and $4_F = 1_F$, and more generally in the field of integers $\mathbb{Z}_p$ for a prime number *p*. However, in an ordered field *F*, the terms $1_F$, $2_F$, …, $n_F$, … are successive sums of positive elements and so every one is positive. They must also be different, for if $m_F = n_F$ where $m_F > n_F$, then $k_F$ is zero where $k_F = m_F - n_F$, contradicting the fact that $k_F$ is positive and not zero.

Once we have found an infinite sequence of different terms $1_F$, $2_F$, …, $n_F$, … in our ordered field *F*, we include their additive inverses $-1_F$, $-2_F$, …, $-n_F$, … the zero element, and the fractions $m_F / n_F$ (for $n_F \neq 0_F$) to get a subfield isomorphic to $\mathbb{Q}$.

Again we can shift a level and regard this subfield as *being* the rational numbers $\mathbb{Q}$ and then it can be said that every ordered field *contains* the rational numbers $\mathbb{Q}$. In particular, this means that we can use the familiar number symbols to represent the rational numbers in any ordered field *F*. This gives every ordered field *F* a specific crystalline substructure in the form of the rational numbers.

## UP TO ISOMORPHISM, THERE IS ONLY ONE COMPLETE ORDERED FIELD

A complete ordered field is an ordered field with the additional axiom of completeness which we will formulate here as:

(C) An increasing sequence $(a_n\}$ bounded above by $L \in F$ tends to a limit $a \in F$ where $a \leq L$.

As we know that an ordered field *F* contains the rationals $\mathbb{Q}$ and so it already contains all finite decimals which are just integers over a power of ten. For any integer $a_0$ and digits $a_1$, $a_2$, …, $a_n$, … all between 0 and 9, the sequence of finite decimals $a_0 \cdot a_1$, $a_0 \cdot a_1 a_2$, …, $a_0 \cdot a_1 \ldots a_n$, … is an increasing sequence in *F*, bounded above by $a_0 + 1$. By the completeness axiom, this has a unique limit $a \leq a_0 + 1$. This unique limit *a* is written as the infinite decimal $a_0 \cdot a_1 \ldots a_n \ldots$. There is some work to do to check that the arithmetic operations on these decimals operate as expected to show that the field *F* contains a subfield isomorphic to the real numbers as infinite decimals. Then we must show that every element $x \in F$ can be written as a decimal expansion.

To find the decimal expansion for *x*, we begin by finding an integer *m* such that $m \leq x < m + 1$. It is not possible to have *x* larger than all integers for if this happened, by completeness, the sequence 1, 2, 3, … would be bounded above by *x* and then it would have a limit $k \in F$ where $k \leq x$. But because *k* is the *limit* of the sequence then, by the definition, for any $\varepsilon > 0$, say $\varepsilon = ½$, all the terms after some *N* lie between $k - \varepsilon$ and $k + \varepsilon$ for all $n > N$. But then, the next term $N + 1$ will lie between them, so

$k-\frac{1}{2}<N+1<k+\frac{1}{2}$, and, in particular, $k-\frac{1}{2}<N+1$, so the next term $N+2$ is now *bigger* than *k*, which is a contradiction. So there must be some integer *q* bigger than *x*. By a similar argument, there must be some integer *p* smaller than *x*. We are now working with the familiar integers and, as we count up from *p* to *q*, we find some integer *m* such that $m \le x < m+1$. By an induction argument on *n*, we can then show successively that *x* lies in the interval:

$$a_0+\frac{a_1}{10}+\ldots+\frac{a_n}{10^n} \le x < a_0+\frac{a_1}{10}+\ldots+\frac{a_n+1}{10^n}$$

and go on to prove that *x* is the real number $a_0 \cdot a_1 \ldots a_n \ldots$ representing the limit

$$a_0+\frac{a_1}{10}+\ldots+\frac{a_n}{10^n}+\ldots$$

For university students in the early part of their course this is a technically difficult proof to write out in full, and even mathematicians find it tedious to write out the detail. But once established and the result is part of the knowledge structure of the research mathematician, it can now be used as a formal algebraic structure—a complete ordered field—linked to the embodiment of the real number line and the arithmetic of decimals. The real numbers $\mathbb{R}$ is the unique complete ordered field that includes the rationals $\mathbb{Q}$ as an ordered subfield, the integers $\mathbb{Z}$ and the natural numbers $\mathbb{N}$.

One thing that the real numbers do *not* contain is an infinitesimal, which is 'arbitrarily small' but not zero. For instance, we might seek a positive infinitesimal $o \in \mathbb{R}$ that is smaller than all positive real numbers, but we cannot have one because $o/2 \in \mathbb{R}$ is positive and smaller.

Various experiences—such as the construction of $\mathbb{R}$ from $\mathbb{Q}$ using Dedekind cuts explained in chapter nine—are often described as 'completing the real number line' by 'filling in the gaps between the rationals'. This introduces a subtle met-before that is widely shared in the community: that there is 'no room' on the number line to fit in any more numbers, and certainly no room for infinitesimals.

The Cantor-Dedekind Axiom stating that the real numbers are order isomorphic to the points on a geometric line, also categorically insists that once the rationals have been 'completed' then this fills up the whole geometric line.

This illustrates how the choice of particular set of axioms gives a framework with particular properties and may be interpreted by the community to say that 'infinitesimals cannot exist'. However, it only shows that they cannot exist *as real number. It* does not mean that they cannot occur in another formal system as we will see in the next chapter.

Mathematicians have control over the axioms that they choose. But then the theorems that are proved are a consequence of that choice. Choose to study a complete ordered field, and the consequence is that there can be no infinitesimals in it. Choose the epsilon-delta definition of continuity of a function, and again certain consequences follow. For instance, this might introduce monsters that are continuous almost everywhere but peculiarly discontinuous elsewhere.

A typical crazy function is given by

$$r(x) = \begin{cases} 1/n \text{ if } x \text{ is the rational number } m/n \text{ in lowest terms} \\ 0 \text{ otherwise,} \end{cases}$$

The graph is then discontinuous at every rational point and continuous at every rational. This is a strange monster. The reason is that, in any interval there are both rationals and irrationals, but in any interval $a \le x \le b$ there are only a finite number of rationals $x = m/n$ whose values exceed a given value of $\varepsilon > 0$. Around any irrational therefore, it is possible to find an interval where all nearby points on the graph are within $\varepsilon$, and round any rational $x = m/n$, for $\varepsilon = 1/(2n)$ there are irrationals where points on the graph are more than $\varepsilon$ away.

The graph is caricatured in figure 12.9, plotted with 'large' points. If smaller points were used to approximate the graph better, then it would be seen that most points cluster around the $x$-axis.

The reader may think that such a bizarre function has no place in the dynamically continuous world we live in. It doesn't. Put simply, this function is not formally continuous in any interval $a \le x \le b$ of *real numbers*, so it would not be expected to be dynamically continuous over any interval. However, it *does* satisfy the formal definition of continuity at every irrational point and it will 'pull flat' in a window centered on an irrational value of $x$, but not in a window centered on a rational.

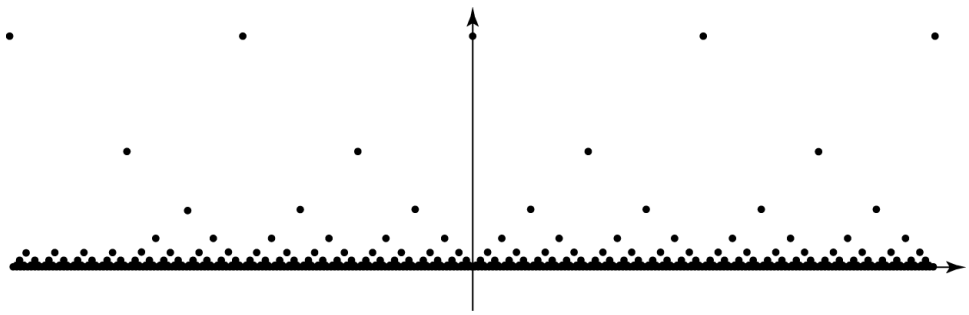

*Figure 12.9: A function continuous at all irrationals and discontinuous at every rational*

What is happening here is that the graph is not formally continuous at *every* real number on an interval, so do not expect it to operate like a dynamically continuous *real* function.

The view expressed by Lakoff and Núñez in *Where Mathematics Comes From* is that many concepts in formal mathematics fail to behave the way that they believe embodied mathematics should operate. They claim that 'romantic' mathematicians produce a form of mathematics that is 'counter-intuitive'. However, in formal mathematics, the mathematician *chooses* his axioms, and the consequences follow. By proving structure theorems mathematicians are able to give axiomatic systems new embodiments that are appropriate for their own prepared knowledge structure. The chosen axioms reveal crystalline formal structures whose consequences are secured by formal proof. These structures may then lead to structure theorems that give previously unimagined embodiments that, for the mathematician, now have a meaningful embodiment.

Different axiomatic systems give different crystalline structures. What is 'natural' depends on the experience of the learner. Learners who have experienced a locally straight introduction to the calculus would not be limited to thinking that 'most' curves are smooth with tangents 'almost everywhere'. They are more likely to *know* that a continuous function may be seriously wrinkled.

The development of axiomatic mathematics is not limited to 'natural' ways of thinking. A mathematician can formulate any consistent system of axioms and deduce theorems to develop a formal knowledge structure, which may then involve structure theorems that give entirely new ways of embodying mathematical structures to a prepared mathematical mind.

Every mathematician began life as a newborn child and has passed through a cognitive development to reach the heights of his mathematical imagination. It is *this* development that we should analyze, from child to mathematician, to find how the embodied brain *becomes* a mathematical mind. Top-down idea analysis from a highly theoretical viewpoint may give some insights, but it is the cognitive growth of the human individual that reveals the true development of human mathematical thinking.

## NEW ORGANIZATIONAL PRINCIPLES

Mathematicians use definitions in a different way from everyday language. In particular, they often take a positive delight in seeing singular examples as part of a general pattern. For instance, the empty set is seen as a central example of a set in its role as a subset of every other set. This violates embodied conceptions where the empty set of apples is evidently different from the empty set of rational numbers. Yet in the axiomatic formal world they are precisely the *same* crystalline concept because they both have the same elements (namely, none).

The empty set is a particular favorite in mathematical formalism. For example, consider the notion of a subspace of a vector space. It is natural to say that the subset of a vector space $V$ (over a field $F$) generated by a subset $S$ consists of all sums $a_1\mathbf{v}_1 + ... + a_n\mathbf{v}_n$ where $a_1$, …, $a_n$ are elements of $F$ and $\mathbf{v}_1,...,\mathbf{v}_n$ are any elements of $S$. If the elements $\mathbf{v}_1,...,\mathbf{v}_n$ are linearly independent then the subspace is said to be $n$-dimensional. For $n = 1$, 2, 3 over the real numbers, these can be seen as one-dimensional, two-dimensional and three-dimensional space. But can the value of $n$ be zero? If the generating subset is the empty set, then how can one have a sum of *no* elements?

This is where the careful use of definition comes into play. The subspace generated by any subset $S$ of a vector space $V$ is defined carefully to be 'the smallest subspace that includes all the elements of $S$.' This includes all the sums $a_1\mathbf{v}_1 + ... + a_n\mathbf{v}_n$ where $\mathbf{v}_1,...,\mathbf{v}_n \in S$, but it must also contain the zero vector, because the rules for a vector space say so. This implies that if $S$ is the empty set, the smallest subspace generated by $S$ must consist of the zero vector alone. This means that a subset generated by $n$ independent vectors has dimension $n$, and this now includes the case $n = 0$.

This subtle generality is the kind of thing that gives great pleasure to mathematicians, even if it seems bizarre in everyday conversation. It is an aesthetically beautiful notion that allows all cases to be included in a general pattern with no exceptions. I remember well as an undergraduate hearing exquisite lectures from the very reflective Oxford don, Ken Gravett. He told us that the empty set was the one that he felt most at home with, because he was absolutely certain what its elements are. If anyone asked, 'is $x$ in the empty set?' (whatever $x$ happened to be), he could always answer 'no.'

Another case of including a singular example as an instance of the general pattern arose when Chris Sangwin asked undergraduates and mathematicians to give examples of even and odd functions.[9] An even function satisfies $f(-x) = f(x)$ and includes functions such as $x^2$ or $\cos x$ and an odd function satisfies $f(-x) = -f(x)$ such as $x^3$ or $\sin x$. The students gave variants of 'typical examples' such as $x^2 + 1$ for even and $\sin 2x$ for odd. A possible response from professional mathematicians, however, was 'zero'. This singular case is both odd *and* even. For a mathematician it is a minimal example, requiring least effort: for a student it is a singular example that fails to be typical for either property.

The philosopher, Paul Grice[10] formulated four maxims for cooperative dialogue, namely quality (truth), quantity (information), relation (relevance) and manner (clarity). The first involves only saying what you believe to be true, for which you have adequate evidence. The second involves making your contribution as informative as possible without giving excessive detail. The third requires what is said to be relevant, the

fourth requires the communication to be brief, avoiding obscurity and ambiguity.

Following Grice's maxims, in everyday language one says things in an informative and suitably clear manner. For instance, one might say 'five is bigger than four', never 'five is greater than or equal to four'. But formal mathematics seeks an economy of means—a parsimony where only what is absolutely necessary is highlighted. This leads to default cases being included in a pattern to minimize the complexity. Mathematicians usually abhor several choices and prefer binary decisions where the response is either yes or no.

For instance, the order property in its strong form $x > y$ requires the use of the trichotomy law: either $x > y$ or $y > x$ or $x = 0$, *and no two hold at the same time.* This follows the Gricean maxim of saying precisely what you know. When I formulated the axioms for order, I framed them using this axiom because I sensed that it would make more sense to readers who are not research mathematicians. It would be less 'natural' to use the order relation $x \geq y$ because in specific arithmetical examples we will always known whether the elements are the same or different.

In the search to reduce every decision as far as possible to a simple yes or no, the notion of order is often given in terms of the corresponding weak order relation $x \geq y$. The weak axioms for order on a set $S$ are:

(WO1) Given $x,\ y \in S$ then either $x \geq y$ or $y \geq x$
(WO2) If $x \geq y$ and $y \geq x$ then $x = y$.

These axioms avoid the three-way test of trichotomy and allow every decision to be binary. Now the definition says that either $x \geq y$ or $y \geq x$ and if both hold then $x = y$.

In mathematics, strong relations (where equality is excluded) are sometimes used in preference to weak relations (where equality is possible). For instance, the notion of parallel lines is a strong equivalence relation between two distinct lines and a line cannot be parallel to itself.

However, in the formal definition of equivalence, an element is always assumed to be equivalent to itself. This is simply a choice made to simplify decision-making. An equivalence relation ~ on a set $A$ is defined to satisfy three axioms:

(E1) $x \sim x$ for all $x \in A$,
(E2) If $x \sim y$ then $y \sim x$,
(E3) If $x \sim y$ and $y \sim z$ then $y \sim z$.

It is quite possible to define a corresponding 'strong equivalence relation' σ defined so that $a\sigma b$ means $a \sim b$ and $a \neq b$. This gives new axioms for a strong equivalence relation in the form:

(SE1) The relation $a\sigma a$ does *not* hold,

(SE2) If $a\sigma b$ then $b\sigma a$,
(SE3) If $a\sigma b$ and $b\sigma c$ where $a \neq c$ then $a\sigma c$.

These axioms lack the simplicity and elegance of (E1)-(E3), and so mathematicians do not use them. Aesthetic sense is part of the axiomatic formal world, choosing axioms and definitions in a form that is elegant and parsimonious, even if this involves different ways of operating from familiar everyday experience.

In some circumstances, changing the list of axioms may give the same underlying crystalline structure, for instance, replacing the axioms for a strong equivalence relation by the more elegant axioms for an equivalence relation, or replacing one version of the completeness axiom in an ordered field by another.

However, when a list of axioms is modified, perhaps by adding or removing an axiom, or making a change in one or more axioms, it often causes a radical change in the crystalline structure. This is not a problem: it is a rich source of freedom, to invent an axiomatic structure that the mathematician chooses for specific purposes. Having chosen a new list of axioms, the consequences of those axioms are not invented, they are discovered.

Often the crystalline structure is so beautiful that even the mathematician who invented the axioms and definitions in the first place is amazed by the sense of perfection that arises. It is no small wonder that a mathematician rich in the experience of discovering beautiful ideas in a given axiomatic system will believe that the crystalline structure is a platonic entity with an existence beyond the confines of his own mind.

Structure theorems translate axiomatic systems and definitions within them into the possibility of new embodiments and new ways of operating symbolically. The power in mathematical thinking lies not just in formulating axioms to describe a familiar situation, it arises from blending together familiar ideas in new ways to solve new problems and, in doing so, to create new crystalline concepts and structure theorems that lead to new forms of embodiment and symbolism underpinned by formal proof.

## References

Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press.

Inglis, M. Mejia-Ramos, J. P. and Simpson, A.P. (2007). Modelling mathematical argumentation: the importance of qualification, *Educational Studies in Mathematics, 66* (7), 3–21.

Mejia-Ramos, J. P. (2008), *The construction and evaluation of arguments in undergraduate mathematics*, PhD Thesis, University of Warwick,

Sangwin, C. J. (2004). Assessing mathematics automatically using computer algebra and the internet. *Teaching Mathematics and its Applications*, 23( 1), 1--14.

Toulmin, S. E. (1958). *The Uses of Argument*. Cambridge University Press.

Weber, K. (2001). Student difficulty in constructing proofs: the need for strategic knowledge. *Educational Studies in Mathematics*, 48 (1), 101-119.

---

[1] Weber (2001).
[2] Weber (2001).
[3] Toulmin (1958).
[4] Toulmin's framework is often presented with the box for the Warrant *below* the line between Data and Qualifier, above the Backing. However, I present it with the Warrant in the main line as I see it as part of the main argument leading to a qualified support for the Conclusion.
[5] Inglis, Mejia-Ramos and Simpson: (2007).
[6] Mejia-Ramos (PhD), 2008
[7] Quoted from Mejia-Ramos (2008), p. 195.
[8] See http://en.wikipedia.org/wiki/Classification_of_finite_simple_groups. Retrieved October 26, 2009.
[9] Sangwin (2004).
[10] Grice (1989).