Diagrammatic Reasoning: Analysis of an Example

P. J. Hayes & G. L. Laforte

IHMC, University of West Florida
11000 University Parkway, Pensacola, Fl. 32514
glaforte,phayes@ai.uwf. edu

Abstract

We argue that a purely diagrammatic proof of a nontrivial mathematical theorem is impossible, because a diagram cannot indicate how it should be generalized. The case is made by subjecting a famous diagrammatic demonstration of Pythagoras' theorem to a close examination, showing that it can be seen to be a demonstration of several different theorems.

Introduction.

We invite the reader, before reading further, to examine figure 1 and consider what, if anything, it suggests to them.

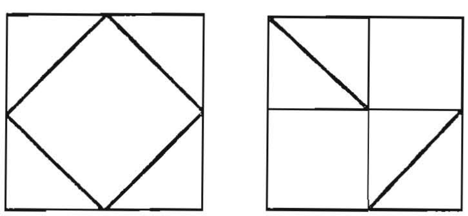

Figure 1

There are many symmetries and coincidences in this figure; many ways, one might say, to understand it. All of the triangles are congruent and symmetric, and could be obtained by bisecting any one of the smaller squares. This figure might be a diagram of a piece of quilting, or an illustration of two tiling patterns, among probably hundreds of other possibilities; it has no obvious mathematical significance. However, if we modify it slightly, it becomes a famous ancient demonstration of Pythagoras' theorem:

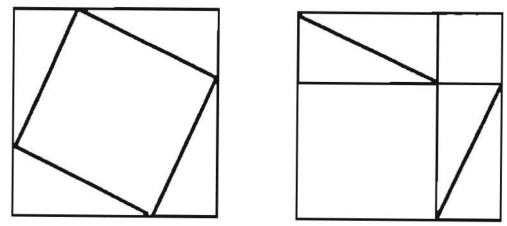

Figure 2

Again, we invite the reader to (in the words of the ancient text) "behold" this figure to see what it suggests. Many people can 'see' this as a demonstration of the truth of Pythagoras' theorem, once they know that is what it is supposed to be.

Some people do not find the demonstration obvious without further explanation. Everyone seems to have a definite `aha!' experience after which the demonstration seems convincingly sound. (For those who have learned that the proof of Pythagoras' theorem is a complex or arcane matter, this moment of realization can provide quite a powerful epiphany; we have witnessed spontaneous gasps or cries of surprise, for example.) Everyone seems to need an initial period of inspection of the diagram before understanding the demonstration, although for many this period is very short. There seems therefore to be some nontrivial cognitive business to be done, even when the purely perceptual machinery has finished (although once the demonstration is familiar, this may be reduced to a simple act of recognition.).

To emphasize this, imagine someone who had no reason to be thinking about Pythagoras' theorem seeing figure 2 casually, without their attention being drawn to it. For them to realize that it constituted a demonstration would be a considerable cognitive step; it would be the kind of effort that we often characterize as a nontrivial act of noticing. Go further, and imagine someone who knew nothing about Pythagoras' theorem and had no special knowledge of areas or right angles, and suppose that person were to realize, upon being shown the diagram, that the theorem was true and that this was a demonstration of it. Such an insight would be an act of genius, the kind that is celebrated in intellectual history, comparable to Archimedes' overflowing bath or Kekuele's snake dream.

The diagram does not contain any information about what it is supposed to be a demonstration of, even assuming that we know that demonstration of some kind is its purpose (which is what the famous instruction "Behold!" may reasonably taken to be, of course.) The viewer needs to interpret the diagram in a particular way in order to perceive the demonstration it illustrates. Contrast this with what we normally speak of as a proof. In normal mathematical text the theorem is stated explicitly as a kind of header to the proof; the final sentence of a formal logical proof is the theorem. In both cases, there is no doubt about what the proof is proving.

(In case the reader does not see why the diagram is a demonstration of Pythagoras' theorem, note that the triangles are all isomorphic copies of the same right-angled triangle, and that each large square contains exactly four copies, occupying the same area in both squares, and therefore leaving the same area unoccupied. In one, this remainder is the (tilted) square on the hypotenuse of the triangle; in the other it has two parts, which are the squares on the other two sides of the triangle.)

In the rest of this paper we will refer to Fig. 2 simply as 'the diagram', to Pythagoras' theorem as 'the theorem', to the use of the diagram to establish the theorem as 'the demonstration', and to the previous paragraph as 'the explanation'.

On generalizing.

"Seeing" the demonstration does not involve literally seeing anything in the diagram that was previously invisible. Some further process of reasoning or comprehension is taking place. Without delving into the psychological details, we can ask what it is that people come to know when they understand the demonstration which they did not know previously.

The above explanation provides a clue, since it is in fact (deliberately) incomplete. It refers only to the particular triangles and squares in the diagram; but the theorem refers to *any* right-angled triangle. The generalization seems so obvious that this step is often missed, but it is crucial to the proof. The diagram is in fact acting as a kind of representative sampler of an infinite class of diagrams, including many which would be impossible to perceive (where the side of the large squares is a light-year but of the smaller square only the diameter of a proton, say.) The step, from seeing this particular diagram to understanding the demonstration, involves the realization that the right-angle triangle displayed there could have been any right-angled triangle, and the explanation would have worked just as well.

The logic of this generalization is similar to the logical rule of universal generalization, which allows us to infer *(Forall x) (P x)* from *(P a)* provided that the name *a* has no conditions placed upon it by any assumptions used in the proof of *(P a)*. The intuitive justification for this apparently counterintuitive rule is that if no particular conditions are placed on the name, then it could equally well have denoted anything at all, so the thing proved true of it must be true of everything. However, this rule cannot be applied directly to the information displayed in the diagram. Only part of it must be generalized, and the diagram itself does not indicate which parts to apply the rule of generalization to.

This is made clearer if one contrasts a description of the diagram to the diagram itself, since a description does suggest a natural generalization. Using a standard technique we might describe the diagram as follows.

There are congruent squares ABCD and EFGH, with a point P between A and B. Points J, K, L, M, N, R and S are located on BC, CD, DA, EF, FG, GH and HE respectively so that $|AP| = |BJ| = |CK| = |DL| = |MF| = |FN| = |GR| = |SE|$, with lines connecting PJ, JK, KL, LP, RN, MR and SN. The last two lines intersect at a point X, and a line joins E to X. The result is shown in figure 3.

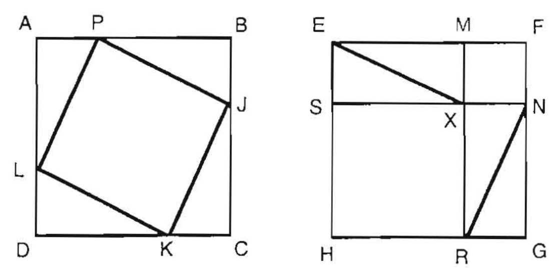

Figure 3

Strictly speaking, this is not the diagram; it is cluttered with labels which confuse the visual field and add nothing to the demonstration. But more significantly, notice that this description does not itself mention triangles at all. It requires a further perceptual step to see the right triangles in the figure. It is essentially a geometric construction algorithm with two free variables (the lengths of $|AB|$ and $|AP|$) so this is more a description of a particular category than a description of the diagram itself. It already has a generalization incorporated into the way the diagram is described; but *this* generalization does not mention triangles.

It is possible to describe the diagram differently, mentioning triangles explicitly, in the following way.

Take a right triangle and make four copies of it, rotating each one clockwise by a right angle with respect to the next. Align the vertices of these triangles to form a large square. Then take a similar square and place inside it four more copies of the triangle, arranged in two rectangles placed into opposite corners of the square.

This description draws attention to the triangles, but leaves out other significant facts (such as that the corners of the two rectangles coincide at X) which now require further inferences to be made. Like the previous description, however, this has a free parameter - the initial right triangle - and is therefore similarly prepared for universal generalization.

A description always mentions only some of the aspects of the structure shown in the diagram. But the descriptions always demonstrate the appropriate generalization, by specifying which aspects of the diagram are arbitrary and therefore susceptible to universal generalization.

This is why figure 1 doesn't work as a demonstration of the theorem, even though, just like the figure, it is simply one special case of the class exemplified by the figure. The extra perceptual coincidences when the right-angled triangle is symmetric distract attention from the intended generalization, and suggest alternative classes of figures – alternative generalizations – that it might be taken to illustrate.

Movies and other generalizations

The demonstration is made more vivid if the triangles are thought of as solid pieces of planar surface, i.e. as triangular tiles which can be moved around. If we think in this way, then the diagram can be reconstructed as follows. Take four copies of a right-triangular tile which fit inside a large square, one in each corner and touching at their tips, so that they fit around a square made from their hypotenuseia. Now slide the leftmost tile down and to the right until its hypotenuse fits against the opposite tile, and then slide the bottom tile upwards, and the top tile to the left, until they meet in a similar way. Figure 4 shows the sequence of movements.

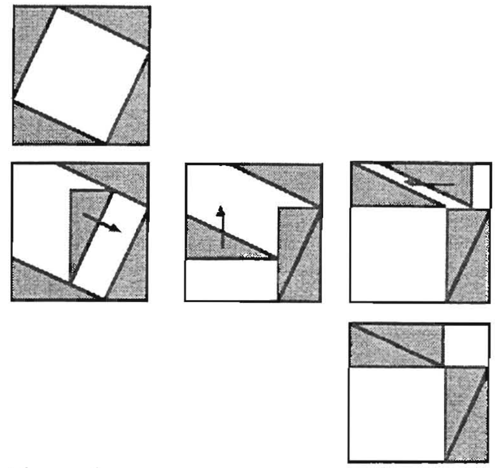

Figure 4

This makes several things very clear: the relationship between the two halves of the diagram, the preservation of area, and the length identities mentioned in the first description. The immediacy of this version of the demonstration depends on the fact that it can be expressed in terms which are so familiar from our everyday physical intuition. We all know that sliding a rigid object preserves its shape and its area, and that a right triangle is exactly half a rectangle; and this knowledge is at the surface of our thinking, as it were. Conclusions about such matters seem to require no effort or search, and require no further explanation or justification. They seem to be simply obvious. But it is important to realize that they are inferences. That our perceptual apparatus is very efficient at drawing such conclusions is a fact about our perception of space, but the conclusions so reached are, indeed, conclusions — they are based on further facts which are not clearly present in the diagram itself.

Just as figure 1 is, while a special case, a member of the same class of diagrams as figure 2, this figure itself is also a special case of a larger class of diagrams. This tile-sliding process works just as well when the triangles involved fit into a parallelogram rather than a square; in fact, they survive any shear, stretch, reflection or rotation applied to the diagram, as illustrated by figure 5.

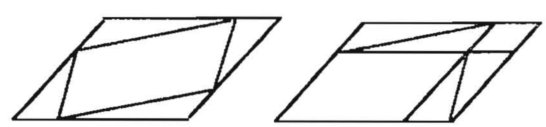

Figure 5

The reason that figure 5 is not considered to be a useful demonstration is not that it fails to demonstrate anything. There is a theorem it demonstrates, and indeed this theorem has Pythagoras' theorem as a special case, just as the theorem suggested by figure 1 is a special case of Pythagoras'. It could be phrased thus: Given two triangles ABC and DEF, if the angles ABC and DEF are complementary, then the sum of the areas of the parallelogram with sides AB and DE, and sides AC and DF, both with angle ABC, is equal to the area of a parallelogram with sides AC and DF and angle (pi - ACB - EDF)

This theorem is not usually considered to be worth stating; but from a purely diagrammatic perspective, its demonstration is just as vivid and direct as the demonstration of Pythagoras: in fact, it is a generalization of that very demonstration.

We have discovered some other generalizations of Pythagoras which also have direct diagrammatic demonstrations using the same sliding-triangular-block method. For example, if one allows the initial triangle to be obtuse rather than right, the demonstration still works, as shown in figure 6. Notice that the two halves of the figure, while not squares, are isomorphic and hence have the same area.

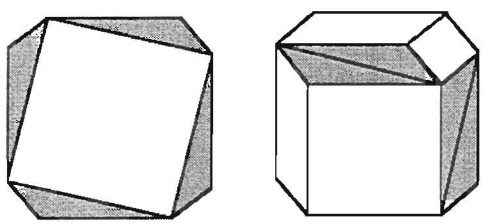

Figure 6

A cursory inspection shows that this demonstrates that the square on the hypotenuse of an obtuse triangle is greater than the sum of the squares on the other two sides; a slightly more careful inspection, together with a small amount of trigonometry, shows it to be a demonstration that the area of the square on the hypotenuse is the sum of the areas of the other sides squared plus twice their product with the cosine of the obtuse angle, i.e.

$$|AC|^2 = |AB|^2+|BC|^2 + 2|AB||AC|\mathrm{Cos}(ABC)$$

We were not previously aware of these generalizations of Pythagoras' theorem, and discovered them by systematically trying various ways to remove the implicit assumptions used in interpreting the usual figure. The step from the diagram in figure 2 to figure 6 seems intuitively peculiar, but this intuition is not itself included in the diagram, or even in the way that the diagram is interpreted and understood. One sees that the outlines in the diagram must be squares if the triangles are right-angles, but this fact, as figure 6 shows, is irrelevant to the demonstration.

Perception and understanding

As we have argued, to understand the demonstration one must choose some aspects of the diagram and generalize over them. Choosing the right aspects is crucial, since the diagram has many ways to be generalized, several of which have interesting mathematical content.

In order to grasp the relevant generalizations, one must perceive the relevant structure in the diagram: the congruent triangles, in particular. In order to compute that these triangles are congruent requires the perceptual process to recognize the equality of their sides. One needs to (literally) see that $|AP| = |CK| = |MX|$, etc., in order to perceive that the triangles are congruent and hence be able to apply universal generalization to the appropriate parts of the diagram.

Note, however, that the particular line lengths involved in these perceptual judgments are precisely those which are lost in the generalization itself. That is, the perceptual process of seeing the diagram must make precise estimates of line lengths, but the generalization must then ignore these lengths. If the acts of visual perception of the diagram – of seeing the lines on the page – and of comprehension of the demonstration – of understanding why this shows that the theorem applies to any right triangle – were unified into a single process of diagrammatic comprehension, then the very involvement of these particular lengths in the machinery of perception would render invalid the generalization on which the demonstration depends. The rule of universal generalization is valid only when no assumptions are made about the things named in its antecedent – the '*a*' in '*P(a)*'. But the perceptual process must make such assumptions: in fact, it must make use of very precise and particular information about the particular line segments displayed in the diagram.

What this seems to show is that the processes of perception and comprehension must be separate, and that the latter depends on constructing a suitable internal *description* of the diagram whose syntactic structure defines which aspects of the diagram can be generalized. The "aha" experience may well be the construction of a suitable such description, one which generalizes naturally to give the well-known theorem.

The assumption of relevance

There seems to be a principle at work here rather similar to one of Grice's conversational maxims: *do not give irrelevant information*. The figure shows us which lines must be identical in length, and it doesn't show us any other accidental coincidences which would be irrelevant to the point it is making. (It is supposed to generalize to these as well, so it should not be understood as denying them.) The theorem is true of symmetric triangles, and figure 1 is then the appropriate case of the figure; but it is not an appropriate exemplar of the class to use in a demonstration (in, one might say, a conversational act of persuasion).

Notice that our hypothetical ignorant genius might well have realized, on seeing figure 1, that it was a proof that the square on the hypotenuse of a symmetric right triangle had twice the area of the square on its short sides, without making the further generalization to Pythagoras' theorem. This would satisfy both the Gricean maxim and normal heuristic rules of generalization, but applied to figure 1 rather than figure 2. This assumption of no irrelevant information seems to be crucial in moving from a diagram to the appropriate description of it to use as a basis for generalization. The reason we do not "see" the diagram as a special case of figure 6 is that if its communicative purpose was to suggest that generalization, then it would have been a very peculiar, and therefore misleading, special case to have chosen. This in spite of the fact that it is indeed a proper case, and that the cosine generalization of the theorem applies to it quite correctly.

Conclusion

We have focused on this famous example in order to make some very general claims. A single diagram cannot be a proof, because any diagram can be described in different ways which yield different generalizations. The process of seeing a diagram involves detailed comparisons of distances which need to be explicitly eliminated by the comprehension process, and this distinction forces us to discuss an intermediate descriptive representation to be appropriately generalized. The process of comprehension seems to involve a Gricean assumption that an illustrative diagram contains no irrelevant coincidences, and therefore any such diagram has meaning only when seen as part of a larger communicative act. Deliberately violating these communicative rules makes even this famous example, well-known for over two millennia, yield surprising new interpretations.

Acknowledgements

This work was supported in part by the National Imagery and Mapping Agency under grant #NMA202-97-1-1020