# Bridging the gap between argumentation theory and the philosophy of mathematics

Alison Pease,[1] Alan Smaill,[1] Simon Colton,[2] John Lee[1]
[1]*School of Informatics, University of Edinburgh*
[2]*Department of Computing, Imperial College London*

**Abstract.** We argue that there are mutually beneficial connections to be made between ideas in argumentation theory and the philosophy of mathematics, and that these connections can be suggested via the process of producing computational models of theories in these domains. We discuss Lakatos's work (1976) in which he championed the informal nature of mathematics, and our computational representation of his theory. In particular, we outline our representation of Cauchy's proof of Euler's conjecture, which uses work by Haggith on argumentation structures, and identify connections between these structures and Lakatos's methods.

**Keywords:** Lakatos, argumentation, philosophy of mathematics, computational model, theory refinement.

## 1. Introduction

Mathematicians have traditionally attributed great importance to proofs, since Euclid's attempts to deduce principles in geometry from a small set of axioms.[1] Experience of the fallibility of proofs led many mathematicians to change their view of the principal role that proof plays from a guarantee of truth to new ideas such as an aid to understanding a theorem (Hardy, 1928), a way of evaluating a theorem by appealing to intuition (Wilder, 1944), or a memory aid (Polya, 1945). Lakatos (1976), via the voice of the teacher, suggested that we see a proof as a thought-experiment which "suggests the decomposition of a conjecture into subconjectures or lemmas, thus *embedding it* in a possibly quite distant body of knowledge" (Lakatos, 1976, p. 9). This change in the perception of the role of proof in mathematics, from its lofty pedestal of infallible knowledge to the more familiar level of flawed and informal thought, suggests that work in argumentation theory, commonly inspired by practical argument such as legal reasoning, may be relevant to the philosophy of mathematics. Conversely, the fertile domain of mathemat-

[1] We are grateful to one of our reviewers for highlighting insufficiencies in certain proofs in Euclid's Elements, in terms of reliance on diagrams and physical constructions which were not formally defined. There is a parallel here with Hilbert's programme, which "glossed over subtle points of reasoning and relied heavily, in some cases, on diagrams which allowed implicit assumptions to be made" (Meikle and Fleuriot, 2003, p. 16). As pointed out by our reviewer, the fact that even such stalwarts as Euclid and Hilbert depended on informal argument strengthens our own argument that there is a place for argumentation theory in mathematics.

ical reasoning can be used to evaluate and extend general argumentation structures.

The relationship between the philosophy of mathematics and argumentation theory has already borne fruit. In Toulmin's well-known model of argumentation (1958), written as a critique of formal logic, he argued that practical arguments focus on justification rather than inference. His layout comprises six interrelated components: a claim (the conclusion of the argument), data (facts we appeal to as the foundation of the claim), warrant (the statement authorising the move from the data to the claim), backing (further reason to believe the warrant), rebuttal (any restrictions placed on the claim), and a qualifier (such as "probably", "certainly" or "necessarily", which expresses the force of the claim). While Toulmin did consider mathematical arguments (for example, Toulmin, 1958, pp. 135–136), he initially developed his layout to describe non-mathematical argument.[2] However, he later applied the layout to Theaetetus's proof that there are exactly five platonic solids (Toulmin et al., 1979). Aberdein has shown that Toulmin's argumentation structure can represent more complex mathematical proofs; such as the proof that there are irrational numbers $\alpha$ and $\beta$ such that $\alpha^\beta$ is rational (Aberdein, 2005), and the classical proof of the Intermediate Value Theorem (Aberdein, 2006). Alcolea (1998) has shown that Toulmin's argumentation structure can also be used to represent meta-level mathematical argument, modelling Zermelo's argument for adopting the axiom of choice in set theory (described in Aberdein, 2005). Alcolea also presents a case study of Appel and Haken's computer assisted (object level) proof of the four colour theorem (Aberdein, 2005, suggests an alternative representation of this theorem, which also uses Toulmin's layout). Aberdein (2006) describes different ways of combining Toulmin's layout, and uses his embedded layout to represent the proof that every natural number greater than one has a prime factorisation.

Lakatos (1976) championed the informal nature of mathematics, presenting a fallibilist approach to mathematics, in which proofs, conjectures and concepts are fluid and open to negotiation. He saw mathematics as an adventure in which, via patterns of analysis, conjectures and proofs can be gradually refined. Lakatos demonstrated his argument by presenting rational reconstructions of the development of Euler's conjecture that for any polyhedron, the number of vertices (V) minus the number of edges (E) plus the number of faces (F) is equal to two, and Cauchy's proof of the conjecture that the limit of any convergent series of continuous functions is itself continuous. He

[2] For instance, Toulmin argues that "mathematical arguments alone seem entirely safe" from time and the flux of change, adding that "this unique character of mathematics is significant. Pure mathematics is possibly the only intellectual activity whose problems and solutions are 'above time'. A mathematical problem is not a quandry; its solution has no time-limit; it involves no steps of substance. As a model argument for formal logicians to analyse, it may be seducingly elegant, but it could hardly be less representative" (Toulmin, 1958, p. 127).

also presented a rational reconstruction of the history of ideas in the philosophy of mathematics. Lakatos's work in the philosophy of mathematics had three major sources of influence: firstly, Hegel's dialectic, in which the *thesis* corresponds to a naïve mathematical conjecture and proof; the *antithesis* to a mathematical counterexample; and the *synthesis* to a refined theorem and proof (described in these terms in Lakatos, 1976, pp. 144–145); secondly, Popper's ideas on the impossibility of certainty in science and the importance of finding anomalies (Lakatos, 1976, p. 139, argued that Hegel and Popper "represent the only fallibilist traditions in modern philosophy, but even they both made the mistake of preserving a privileged infallible status for mathematics"); and thirdly, Polya's (1954) work on mathematical heuristic and study of rules of discovery and invention, in particular defining an initial problem and finding a conjecture to develop (Lakatos, 1976, p. 7, note 1, claimed that the discussion in his (1976) starts where Polya's stops). Lakatos held an essentially optimistic view of mathematics, seeing the process of mathematical discovery in a rationalist light. He challenged Popper's view that philosophers can form theories about how to evaluate conjectures, but not how to generate them[3] in two ways: *(i)* he argued that there *is* a logic of discovery, the process of generating conjectures and proof ideas or sketches *is* subject to rational laws; and *(ii)* he argued that the distinction between discovery and justification is misleading as each affects the other; *i.e.*, the way in which we discover a conjecture affects our proof (justification) of it, and proof ideas affect what it is that we are trying to prove (see Larvor, 1998). This happens to such an extent that the boundaries of each are blurred.[4]

[3] Popper (1959) argued that the question of generation should be left to psychologists and sociologists: "The question of how it happens that a new idea occurs to man... may be of great interest to empirical psychology; but it is irrelevant to the logical analysis of scientific knowledge" (*ibid.*, p. 31), and shortly after emphasised again, that: "there is no such thing as a logical method of having new ideas" (*ibid.*, p. 32).

[4] The question of whether mathematical claims and proofs are socially constructed is clearly pertinent to our own thesis that argumentation theory is relevant in a mathematical context. Whether Lakatos held a social-constructivist philosophy of mathematics is controversial (but perhaps irrelevant for our current thesis). However, there are certainly social aspects in Lakatos's theory: in particular his emphasis on the influence that Hegel's dialectic had on his thinking, and his presentation of mathematical development as a social process of concept, conjecture and proof refinement, presented in dialogue form. Ernest (1997) develops Lakatos's fallibilism and ideas on negotiation and acceptance of mathematical concepts, conjectures and proofs (together with Wittgenstein's socially situated linguistic practices, rules and conventions) into a social-constructivist philosophy of mathematics. Goguen (1999) also presents a defence of the social-constructivist position, arguing that although mathematicians talk of proofs as real things, "all we can ever actually find in the real world of actual experience are proof events, or "provings", each of which is a social interaction occurring at a particular time and place, involving particular people, who have particular skills as members of an appropriate mathematical social community" (*ibid.*, p. 288). He continues in a very Lakatosian vein to

There are a few explicit and implicit overlaps between Lakatos's work and argumentation theories. Aberdein (2006) uses Toulmin's layout to describe and extend Lakatos's method of lemma-incorporation, where a rebuttal ($R$) to a claim ($C$) is a global counterexample, and the argument is repaired by adding a lemma ($L$), which the counterexample refutes, as a new item of data ($R \rightarrow \neg L$) and incorporating the lemma into the claim as a precondition ($L \rightarrow C$). Pedemonte (2000) discusses the relationship between the production of a mathematical conjecture and the construction of its proof-object, referred to as *cognitive unity*. This can be compared to Lakatos's *logic of discovery and justification*. While Lakatos viewed the two processes as circular, with changes in the proof suggesting changes to the conjecture and vice versa, as opposed to Pedemonte's assumption of a chronological order, the two theories can be seen as cognitive and philosophical counterparts. Naess's (1953) work on argumentation can also be compared to Lakatos's theory: he argues that discussion can be about interpretation of terms, during which a process of precization takes place. If this fails to lead to agreement then evidence is weighed up to see which of two interpretations is more acceptable. There are very strong parallels between this and Lakatos's method of monster-barring in particular, but also his other methods of conjecture and proof refinement. Another similarity is the *degree* of precization required: both Lakatos and Naess (1966) argued that we should make our expressions sufficiently precise for the purposes at hand, rather than aim to resolve all ambiguity.[5]

The rest of this paper is structured as follows: in §2 we describe our computational model of Lakatos's theory, HRL, which has suggested new connections between argumentation theory and the philosophy of mathematics. In §3 we discuss how we have represented Cauchy's proof of Euler's conjecture by using work by Haggith on argumentation representation and structures. §4 contains a discussion of aspects of Lakatos's method of lemma-incorporation and how they have affected our algorithmic realisation: we also describe our algorithms for each type of lemma-incorporation and for determining which type to perform. In §5 we outline some connections between Haggith's argumentation structures and Lakatos's methods and show how other mathematical examples can be described in this way. We conclude in §6.

criticise the way mathematicians often hide obstacles and difficulties when presenting their proofs, advocating that the drama be reintroduced.

[5] We thank one of our reviewers for pointing out the relevance of precization. $U$ is more precise than $T$ if any interpretations of $U$ are also interpretations of $T$, but there are interpretations of $T$ which are not interpretations of $U$. Both Naess and Lakatos see agreeing on the meaning of terms as a stage of a discussion (as opposed to a pre-requisite to it as, for example, in Crawshay-Williams, 1957). Naess also shares with Lakatos an approach to philosophy which is mainly based on descriptive, rather than normative aspects.

## 2. A computational model of Lakatos's theory

Lakatos outlined various methods by which mathematical discovery and justification can occur. These methods suggest ways in which concepts, conjectures and proofs gradually evolve via interaction between mathematicians, and include surrender, monster-barring, exception-barring, monster-adjusting, lemma-incorporation, and proofs and refutations. Of these, the three main methods of theorem formation are monster-barring, exception-barring, and the method of proofs and refutations (Lakatos, 1976, p. 83). Crudely speaking, monster-barring is concerned with concept development, exception-barring with conjecture development, and the method of proofs and refutations with proof development. However, these are not independent processes; much of Lakatos's work stressed the interdependence of these three aspects of theory formation. We hypothesise that *(i)* it is possible to provide a computational reading of Lakatos's theory, and *(ii)* it is useful to do so. To test these two hypotheses we have developed a computational model of Lakatos's theory, HRL.[6] Running the model has provided a means of testing hypotheses about the methods; for instance that they generalise to scientific thinking, or that one method is more useful than another. Additionally, the process of having to write an algorithm for the methods has forced us to interpret, clarify and extend Lakatos's theory, for instance identifying areas in which he was vague, or omitted details.[7]

In keeping with the dialectical aspect of (Lakatos, 1976), our model is a multiagent dialogue system, consisting of a number of student agents and a teacher agent. Each agent has a copy of the theory formation system, HR (Colton, 2002), which starts with objects of interest (*e.g.*, integers) and initial concepts (*e.g.*, division, multiplication and addition) and uses production rules to transform either one or two existing concepts into new ones. HR also makes conjectures which empirically hold for the objects of interest supplied. Distributing the objects of interest between agents means that they form different theories, which they communicate to each other. Agents then find counterexamples and use methods identified by Lakatos to suggest modifications to conjectures, concept definitions and proofs.

In this paper we are concerned with our algorithmic realisation of the method of proofs and refutations, and its mutually beneficial association with

[6] The name incorporates HR (Colton, 2002), which is a system named after mathematicians Godfrey Harold Hardy and Srinivasa Aiyangar Ramanujan and forms a key part of our model, and the letter "L", which reflects the deep influence of Lakatos's work on our model.

[7] Since Lakatos's work (1976) was the first attempt to characterise informal mathematics (see Corfield, 1997 and Feferman, 1978), it is likely to be incomplete, and hence be open to criticism and extension. Lakatos himself neither considered the methods complete nor definitive, arguing only that they provide a more realistic and helpful portrayal of mathematical discovery than Euclidean (deductive) methodology.

work from argumentation theory. Other aspects of the project are described in (Pease et al., 2004).

## 3. A computational representation of Cauchy's proof

The method of *lemma-incorporation*, developed via the dialectic into the method of *proofs and refutations*, is considered by Lakatos to be the most sophisticated in his (1976). Commentators and critics, for instance Corfield (1997) or Feferman (1978), usually share this view, often seeing the rest of the book as a prelude to this method.[8] The method works on a putative proof of a conjecture: the main example in (Lakatos, 1976) is Cauchy's (1813) proof of Euler's conjecture that for all polyhedra, $V - E + F$ is 2. Therefore, in order to model *lemma-incorporation* in HRL, we need a way of representing an informal mathematical proof.

### 3.1. Cauchy's proof

Lakatos argued that nineteenth century mathematicians viewed Cauchy's proof of Euler's conjecture in (Cauchy, 1813) as establishing the truth of the 'theorem' beyond doubt (Lakatos, 1976, p. 8, cites Crelle, 1827, Jonquières, 1890, and Matthiessen, 1863 as examples). For a diagrammatic representation of these steps, carried out on the cube, see Figure 1, taken from (Lakatos, 1976, p. 8).

> *Step 1:* Let us imagine the polyhedron to be hollow, with a surface made of thin rubber. If we cut out one of the faces, we can stretch the remaining surfaces flat on the blackboard, without tearing it. The faces and edges will be deformed, the edges may become curved, but $V$ and $E$ will not alter, so that if and only if $V - E + F = 2$ for the original polyhedron, $V - E + F = 1$ for this flat network — remember that we have removed one face. *Step 2:* Now we triangulate our map — it does indeed look like a geographical map. We draw (possibly curvilinear) diagonals in those (possibly curvilinear) polygons which are not already (possibly curvilinear) triangles. By drawing each diagonal we increase both $E$ and $F$ by one, so that the total $V - E + F$ will not be altered. *Step 3:* From the triangulated map we now remove the triangles one by one. To remove a triangle we either remove an edge — upon which one face and one edge disappear, or we remove two edges and a vertex — upon which one face, two edges and one vertex disappear. Thus, if we had $V - E + F = 1$ before a triangle is removed, it remains so after the triangle is removed. At the

[8] This is possibly because, despite different perspectives on the role of proof in mathematics, the idea that it is an important one is generally accepted, and these are the only methods to consider the 'proof' of a conjecture.

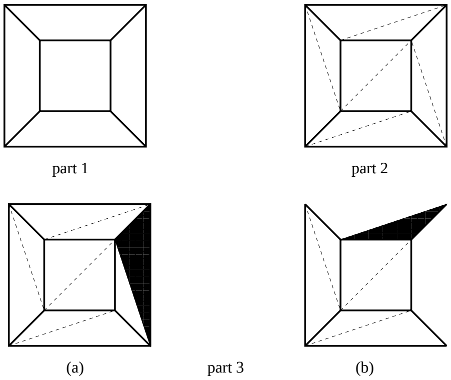


*Figure 1.* Given the cube, after removing a face and stretching it flat, we are left with the network in part 1. After triangulating, we get part 2. When removing a triangle, we either remove one edge and one face, or two edges, one vertex and a face — shown in parts 3(a) and (b) respectively.

> end of this procedure we get a single triangle. For this $V - E + F = 1$ holds true. (Lakatos, 1976, pp. 7–8)

### 3.2. Representing informal mathematical proofs

Cauchy's proof can be represented in a variety of ways using Toulmin's layout. We show one of the simplest ways in Figure 2; more sophisticated versions might include multiple-linked, or nested layouts (as in Aberdein 2005, 2006) where individual proof steps each form a *claim* in one argument, which then (possibly combined with other premises) forms the *data* in a subsequent argument.[9] However, while Toulmin picked apart argumentation structures and showed how the traditional "Minor Premise, Major Premise, so Conclusion" was too crude to represent the way in which people actually argue, he mainly identified different types of statement which in some way *support* a claim. There is only one type of statement in Toulmin's layout which *opposes* a claim: a rebuttal. This can be interpreted in different ways: as a rebuttal to the claim, a rebuttal to the warrant, a rebuttal to an implicit premise, or as a statement which supports a refutation of the claim, and its function is still under debate (see Reed and Rowe, 2005, pp. 15–19). Pollock (1995) defines a *rebutting defeater* $R$ as a reason for denying a claim $P$ which is supported by prima facie reason $Q$. He claims (*ibid.*, p. 41) to have been the first to explicitly point out defeaters other than a rebuttal, in (Pollock, 1970), and identifies the *undercutting defeater*. This defeater attacks the connection between a prima facie reason and the conclusion, rather than attacking the

[9] It would be interesting to investigate how many of the arguments described in (Lakatos, 1976) can be represented in this layout.

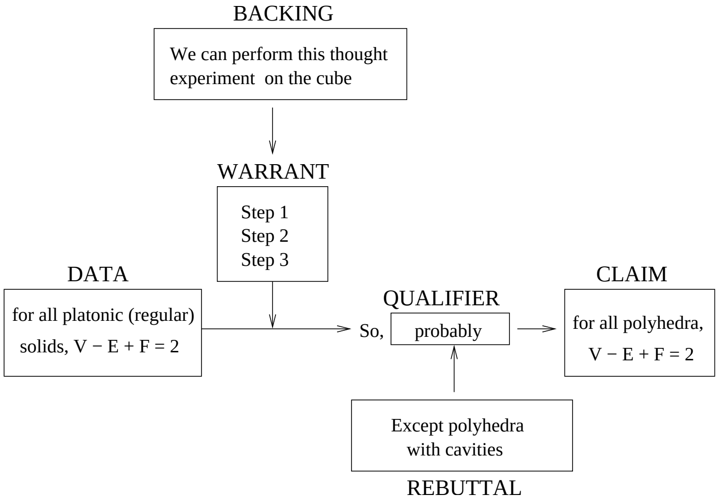

BACKING
We can perform this thought experiment on the cube
WARRANT
Step 1
Step 2
Step 3
DATA
for all platonic (regular) solids, V − E + F = 2
So,
QUALIFIER
probably
CLAIM
for all polyhedra, V − E + F = 2
Except polyhedra with cavities
REBUTTAL

*Figure 2.* A representation of Cauchy's proof of Euler's conjecture, using Toulmin's layout, where steps 1–3 are as described above, taking into consideration the first counterexample. The data are the facts which initially inspire the conjecture in Lakatos's 1976.

conclusion directly. For the purposes of our computational model we have adopted a meta-level argumentation framework (Haggith, 1996), consisting of a catalogue of argument structures which give a very fine-grained representation of arguments, in which both arguments and counter-arguments can be represented. While this framework may lack Toulmin's analysis of statements which support a claim (and has no way of representing a qualifier), it is clear which part of an argument a rebutter rebuts. Given Lakatos's emphasis on both the importance and the different types of counterexample, we deemed this framework appropriate for our needs.

### 3.3. Haggith's argumentation structures

Haggith (1996) starts from the viewpoint that if a domain is controversial, then there may be more than one answer to a question and therefore disagreements may be useful, rather than an obstacle to be overcome. The primary goal of the system described by Haggith, therefore, is to *explore* rather than resolve conflicts. In order to incorporate a high degree of flexibility, Haggith represents arguments at the meta-level which is independent of logic or any specific representation language or domain.

Haggith's representation language describes three categories of meta-level object: proposition names, arguments and sets. The symbols used in the alphabet are: $A_1$, $A_2$, ... to denote proposition names, "$\Leftarrow$" the argument constructor, the standard set and logical symbols, some relation names (such as disagree) and brackets and commas. An argument $A$ in which $C$ is the conclusion, derived from premises $P1$ and $P2$, where $P2$ is itself derived from premise $P3$, is represented as $A = \{C \Leftarrow \{P1, P2 \Leftarrow \{P3\}\}\}$. There are two destructor relations for looking inside argument and set terms: *set membership*, $\in$, which is a two-place, infix relation where $P \in S$ if, for some $S1$, $S = S1 \cup \{P\}$; and *argument* which is a three-place relation, where argument$(A, P, S)$ if $A$ is the argument $P \Leftarrow S$ and $P$ is called the conclusion and $S$ is called the premise set. A further destructor, the support relation can be defined from the other two. This holds between two propositions, the second of which occurs in an argument for the first: supports$(P, Q)$ holds if there exists $A$, argument$(A, P, S)$ and, either $Q$ is a member of $S$, or there exists $A1$, a member of $S$, such that argument$(A1, P1, S1)$ and supports$(P1, Q)$. Haggith has defined four primitive relations at the meta-level which express links and contrasts between object level propositions: $equivalent(P, Q)$, where $P$ and $Q$ are names of propositions which mean the same; $disagreement(P, Q)$, where $P$ and $Q$ are names of propositions which disagree or express a conflict; $elaboration(P, S)$, where $P$ is a proposition name and $S$ is a set of names of propositions which elaborate or embellish upon $P$; and $justification(P, S)$, where $P$ is a proposition name and $S$ is a set of names of propositions which are a justification of $P$. Haggith provides some properties of these relations which restrain their possible applicability, for instance: $disagreement(P, Q) \rightarrow disagreement(Q, P)$; $disagreement(P, Q) \rightarrow not(equivalent(P, Q))$; and $(equivalent(P, Q)$ & $elaboration(P, S)) \rightarrow elaboration(Q, S)$.

Haggith then constructs higher order, meta-level relations defined in terms of the four primitive relations. It is Haggith's development of these argumentation structures that distinguishes her work from standard box-arrow systems. We give two of the structures in detail here and sketch the rest below. We use the letters "$X$" and "$Y$" to represent anonymous variables. *Rebuttal*, inspired by (Elvang-Gøransson et al., 1993), is a relation between arguments whose conclusions disagree. The meta-level definition is:

- $rebuttal(P)$ is the set of arguments, $A$, such that $disagreement(P, Q)$ & $argument(A, Q, X)$.
- $rebuts(A, B)$ if $argument(A, P, X)$ & $member(B, rebuttal(P))$.

That is, the rebuttal of a proposition $P$ is the set of arguments for any propositions which disagree with $P$. Two arguments rebut each other if one is a member of the rebuttal of the conclusion of the other. Haggith's notion of rebuttal fits with an interpretation of the Toulminian rebutter as one which rebuts the claim. It also coincides with Pollock's definition of rebuttal.

*Undercutting* is inspired by (Toulmin, 1958) and defined as follows: an argument $A$, with conclusion $Q$, undercuts an argument $B$ if for some premise $P$ of $B$, $P$ disagrees with $Q$. That is, an argument undercuts another, if the first rebuts a premise of the second. The meta-level definition is:

- $undercutting(P)$ is the set of arguments, $A$, such that $argument(X, P, S)$ & $member(P1, S)$ & $disagreement(P1, Q)$ & $argument(A, Q, Y)$.[10]
- $undercuts(A, B)$ if $argument(A, P, X)$ & $member(B, undercutting(P))$.

Haggith's notion of undercutting fits with an interpretation of the Toulminian rebutter as one which rebuts the warrant, and with Pollock's definition of undercutting. The only difference is that both the interpretation of Toulmin and Pollock specify the *type* of supporting premise which a statement must rebut (the warrant and the prima facie, respectively), while Haggith does not make that distinction.

Given a proposition *P* and an argument *A* for *P*, possible argument moves which provide support for *P* include:

- *corroboration*: an argument for a proposition which is equivalent to (or is) *P*;
- *enlargement*: an argument for an elaboration of *P*, and
- *consequence*: an argument in which *P* is a premise.

Argument moves which oppose *A* include:

- *rebuttal*: an argument for a proposition which disagrees with *P*;
- *undermining*: an argument for a proposition which disagrees with a proposition which is an elaboration of, or is equivalent to, *P*;
- *undercutting*: an argument for a proposition which disagrees with a premise of *P*;
- *target*: an argument which contains a premise which disagrees with *P*, and
- *counter-consequence*: an argument which contains a premise which disagrees with the conclusion of another argument in which *P* is a premise (inspired by Sartor, 1993).

### 3.4. Using Haggith's argumentation structures to represent mathematical proofs

We have expressed Cauchy's proof in Haggith's terms by writing it as a series of propositions and showing the relationships between them. This is shown in Figure 3, where the proof looks as follows:

$$A = \{C \Leftarrow \{P0, P1 \Leftarrow \{P2 \Leftarrow \{P7\}, P3, P4 \Leftarrow \{P8\}, P5, P6\}\}\}$$

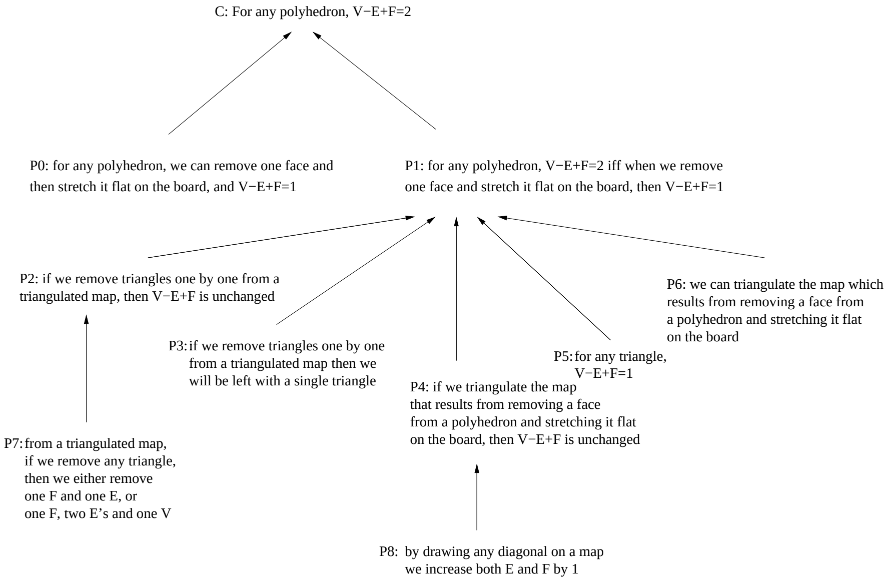

C: For any polyhedron, V−E+F=2
P0: for any polyhedron, we can remove one face and then stretch it flat on the board, and V−E+F=1
P1: for any polyhedron, V−E+F=2 iff when we remove one face and stretch it flat on the board, then V−E+F=1
P2: if we remove triangles one by one from a triangulated map, then V−E+F is unchanged
P3: if we remove triangles one by one from a triangulated map then we will be left with a single triangle
P4: if we triangulate the map that results from removing a face from a polyhedron and stretching it flat on the board, then V−E+F is unchanged
P5: for any triangle, V−E+F=1
P6: we can triangulate the map which results from removing a face from a polyhedron and stretching it flat on the board
P7: from a triangulated map, if we remove any triangle, then we either remove one F and one E, or one F, two E's and one V
P8: by drawing any diagonal on a map we increase both E and F by 1

*Figure 3.* The original proof of Euler's conjecture, represented in Haggith's terms. The arrows represent the justification relation, where the set of premises taken together on any line supports the proposition directly above it. Propositions without any arrows leading into them are unsupported assumptions (thus particularly open to counter-argument).

.

The four initial counter-arguments (all questioning unsupported assumptions), suggested by the ingenious students, to the three steps of Cauchy's proof on (Lakatos, 1976, p. 7) are all examples of Haggith's *target* arguments, disagreeing with a premise of $C$. We represent them below,[11] and show the diagrammatic representation of the first in Figure 4:

1) $-P0$: Some polyhedra, after having a face removed, cannot be stretched flat on a board (questioning the first step).

2) $-P8$: In triangulating the map, we will not always get a face for every new edge (questioning the second step).

3) $-P7$: There are more than two alternatives, when we remove the triangles one by one, that either one edge and a face; or two edges, a face and a vertex disappear (questioning the third step).

4) $-P3$: If we remove triangles one by one from a triangulated map, then we may not be left with a single triangle (also questioning the third step).

[10] Another way of saying this is $argument(X, P, S)$ & $member(P1, S)$ & $member(A, rebuttal(P1))$.

[11] We state the counter-arguments as propositions, whereas in (Lakatos, 1976) they are questions, *i.e.*, "are you sure that..." rather than "it is not possible...".

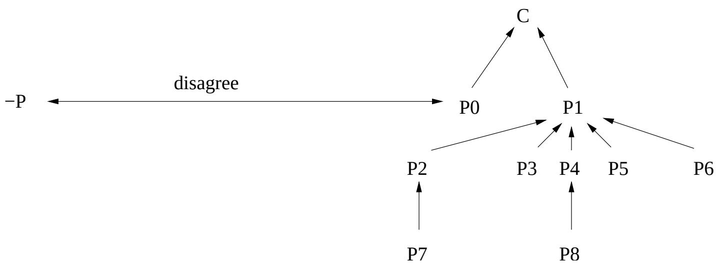


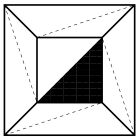
*Figure 4.* The first counter-argument, represented in Haggith's terms. Unmarked arrows represent the justification relation.

## 4. Lakatos's method of lemma-incorporation

Lakatos's method of lemma-incorporation distinguishes *global* and *local* counterexamples, which refute the main conjecture or one of the proof steps (or lemmas), respectively. When a counterexample is found, lemma-incorporation is performed by determining which type of counterexample it is: if it is local but not global (the conclusion may still be correct but the reasons for believing it are flawed) then he proposes modifying the problematic proof step but leaving the conjecture unchanged; if it is both global and local (there is a problem both with the argument and the conclusion) then Lakatos proposes modifying the conjecture by incorporating the problematic proof step as a condition; and if it is global but not local (there is a problem with the conclusion but no obvious flaw in the reasoning which led to the conclusion) then he proposes looking for a hidden assumption in the proof step, then modifying the proof and the conjecture by making the assumption an explicit condition. The method of proofs and refutations consists of setting out to prove and refute a conjecture, looking for counterexamples both to the conjecture and the lemmas, determining which type of counterexample it is, and then performing lemma-incorporation. In the discussion below, we follow Lakatos's convention of using students with names from the Greek alphabet to present different mathematical and philosophical viewpoints.

### 4.1. Three types of counterexample

The cube is a local but not global counterexample since it violates the third lemma in Cauchy's proof, but not the conjecture. That is, it is possible to remove a triangle without causing the disappearance of one edge or else of two edges and a vertex, by removing one of the inner triangles (see Figure 5); in this case, we remove a face but no edges or vertices. In this case we want to modify the proof but leave the conjecture unchanged. We do this by generalising from a single counterexample to a class of counterexamples, and modifying the problem lemma to exclude that class. In this example,



part 3 (c)

*Figure 5.* Given the network which results from taking the cube, removing a face and stretching it flat, and triangulating, we can remove a triangle (shown in black) which results in removing one face, no edges and no vertices.

lemma three becomes "when one drops the *boundary* triangles one by one, there are only two alternatives – the disappearance of one edge or else of two edges and a vertex". The corresponding method is local, but not global lemma-incorporation.

The hollow cube (a cube with a cube shaped hole in it), is both a global counterexample, since $V - E + F = 16 - 24 + 12 = 4$, and local, since it cannot be stretched flat on the blackboard having had a face removed. In this case we need to identify the faulty lemma, lemma one, and then make that step a condition of the conjecture. The proof is left unchanged. Given the hollow cube, we should incorporate the first lemma into the conjecture; this then becomes "for any polyhedron which, by removing one face can be stretched flat onto a blackboard, $V - E + F = 2$". The corresponding method is global and local lemma-incorporation.

The cylinder is a global counterexample, as $V - E + F = 0 - 2 + 3 = 1$ but not local, since it does not violate any of the proof steps. We can remove a face and stretch it flat, resulting in two circles which are either disjoint or concentric (see Figure 6). In order to falsify the second lemma, we would have to draw an edge which joins two non-adjacent vertices, but does not create a new face. Clearly we cannot do this as there are no vertices on the map. Similarly, in order to falsify the third lemma, we would have to be able to remove a triangle and not remove either one edge and one face, or two edges, a vertex and a face, and since there are no triangles on the map, we cannot fail at this stage either. Thus it suggests that the conjecture is flawed and yet the proof of it is upheld. Lakatos argues that this strange situation occurs because of 'hidden assumptions' in the proof, which the counterexample *does* violate. Faced with such a counterexample, he suggests that we retrace our progress through the proof, until we come to a step which is in some way surprising, *i.e.*, one which violates some hidden assumption in the mathematician's mind. Once this has been identified, we should make it explicit in the proof. The counterexample then becomes one of the second type. For instance, one hidden assumption is that having performed lemma one, we are left with a connected network. Therefore we should add this into the proof explicitly, and modify lemma one to ‘any polyhedron, after having a face removed, can be stretched flat on the blackboard, and the result is a connected network’. The cylinder clearly violates this, so we incorporate the new, explicit lemma, into the conjecture statement, which then becomes “for any polyhedron which, after having a face removed, can be stretched flat on the blackboard leaving a connected network, $V - E + F = 2$”. The principle of turning a counterexample which is global but not local into one which is both is called the *principle of retransmission of falsity* in (Lakatos, 1976). This requires that falsehood should be retransmitted from a global conjecture to the local lemmas. Thus, any entity which is a counterexample to the conjecture must also be a counterexample to one of the lemmas. Lakatos called this *hidden lemma-incorporation*.[12]

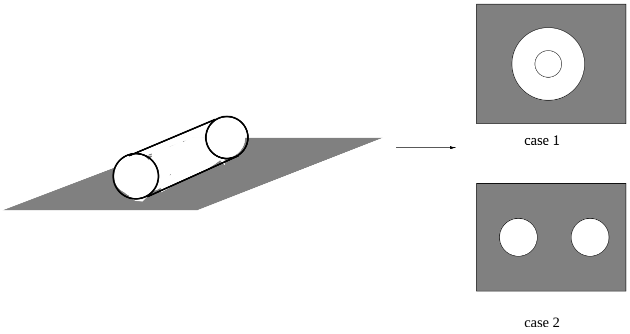


*Figure 6.* If we remove a face from the cylinder and stretch it flat, then we either get case 1 if we remove an end face, or case 2 if we remove the jacket. Either way, we have satisfied the first lemma.

[12] Note that even if we disregard the different interpretations of the second lemma, and hence disagreement about whether the cylinder is a local as well as global counterexample, *Gamma’s* argument that it is only global is not convincing. In the initial proof given (see §3.1) it *does* say explicitly that at the end of the process there is a single triangle: *if we drop the triangles one by one from a triangulated map, we will end up with a single triangle*. This lemma is violated by the cylinder, making it both a global and local counterexample. This would allow for the usual modification of making the lemma a precondition, *i.e.*, the conjecture would become: ‘for any polyhedron which, after having a face removed, and then stretched flat, triangulated and the triangles removed one by one, *leaves a single remaining triangle*, $V - E + F = 2$’.

*Gamma* is able to make his argument because the students get distracted by his claim that the cylinder can be triangulated. The discussion then turns to the meaning of statements which are vacuously true. If they had not disputed this point, *Gamma* would not have been able to uphold his argument. However, the cylinder is still an important example, in that it highlights hidden assumptions in the proof, which should be explicit.

### 4.2. Discussion of lemma-incorporation

In this section, we discuss various aspects of lemma-incorporation and how they have affected our algorithmic realisation.

**Combining the methods**

Exception-barring can be used in lemma-incorporation to get a 'very fine delineation of the prohibited area' (Lakatos, 1976, p 37). This means that we were able to reuse our exception-barring algorithms within our computational representation of the method of lemma-incorporation.

**The type of entity in hidden lemma-incorporation**

The *type* of entity that we are discussing is important for computational purposes. This may change as we step through a proof. For instance, Cauchy's proof begins by referring to *polyhedra* and, once a face has been removed and a polyhedron stretched onto a board, then discusses *graphs*. Since a polyhedron can only be a counterexample to conjectures about polyhedra, not to conjectures about graphs, it may appear that we have a counterexample which is global and not local. It is necessary to look for the corresponding graph and determine whether this entity is a counterexample to those conjectures about graphs. In this case, the disconnected circles in lemma two correspond to the cylinder. This is glossed over in (Lakatos, 1976), as the following quote shows:

> *Gamma:* The cylinder *can* be pumped into a ball — so according to *your* interpretation it does comply with the first lemma.
>
> *Alpha:* Well... But you have to agree that it does *not* satisfy the *second* lemma, namely that '*any face dissected by a diagonal fall into two pieces*'. How will you triangulate the circle or the jacket? Are these faces simply connected? (Lakatos, 1976, p. 44).

When *Alpha* uses the word 'it', he refers to the cylinder. However, he then moves on to talking about the associated graph. While for humans this leap may be acceptable, when implementing this in a program we need to be explicit about the types of entity to which we are referring.

**Identifying a problem lemma in hidden lemma-incorporation**

In a proof where the lemmas chain together as a sequence of implications, the problem of identifying lemmas involving hidden assumptions, as presented by Lakatos, is not difficult for humans. This is because of the element of surprise which people feel when an entity does not "behave" in the expected way, where the 'expected way' has been learned from previous examples. Modelling this feeling of surprise, however, is a difficult task. To help us, we considered what caused the surprise and produced a simple model of that.[13] In Lakatos's example, hidden assumptions are found in two lemmas and cause surprise in different ways.

Lemma one states that any polyhedron, after having a face removed, can be stretched flat onto a blackboard. Although the cylinder is a supporting example of this conjecture, it is surprising: when we remove the jacket from the cylinder, it falls into two parts, leaving two disconnected circles. This is surprising since all previous examples resulted in connected networks. Therefore, we needed to capture the idea of an entity being surprising *with respect to a given conjecture*. We have defined this as follows:

- surprisingness (type 1): an entity $m$ which is a global counterexample to a proposition is surprising with respect to one of the conjectures (or lemmas) $C = \forall x(P(x) \rightarrow Q(x))$ in the proof of the proposition, if another conjecture $C'$ can be found, of the form $\forall x(P(x) \rightarrow (Q(x) \wedge R(x)))$, for some concept $R$, where $m$ is the only known counterexample to $C'$.

Given a proof-scheme and an entity which is a global but not local counterexample, our algorithm for surprise caused by unexpected behaviour is to go through each lemma in the proof-scheme and, if possible, generate a further conjecture $C'$ of the form above. In order to identify the 'hidden assumption' in a conjecture, we have to break down the concepts in it, in particular the concept $Q$ in the conjecture $P \rightarrow Q$. This is made easy for our purposes since for each of its concepts, HR records the construction path, and in particular the concepts to which production rules were applied to get a current concept. This ancestor list allows HRL to gradually dissect a concept until a suitable further concept, $R$, is found.

The discussion of the second lemma, that *any face dissected by a diagonal falls into two pieces*, with respect to the two disconnected circles, is related to work on meaning and denotation (for instance, Russell, 1971, pp. 496–504). The problem is that although there are no diagonals on a circle, we are making a claim about the properties that they have. Lakatos's character *Gamma* argues that it is correct to say that 'every new diagonal we draw on two disconnected circles results in a new face' ($P$), since the negation, that 'there is a diagonal of the two circles which does *not* create a new face' ($\neg P$), is false. This argument uses the law of excluded middle in classical logic, $P \vee \neg P$, *i.e.*, $\neg(\neg P) \rightarrow P$. According to this argument, the cylinder is a global but not local counterexample. *Alpha* disagrees, arguing that if we say that $P$ is true then we must be able to construct at least one instance of it, *i.e.*, there must be an existential clause in the lemma. The statement 'a face

[13] Note that we would not claim that our model itself is surprised, simply that the model can identify those lemmas which cause surprise to humans, and the hidden assumptions within the lemmas.

is simply connected' means 'for all x, if x is diagonal then x cuts the face into two; *and there is at least one x that is a diagonal*' (Lakatos, 1976, p. 45). Under *Alpha*'s interpretation, the cylinder *is* a counterexample to this lemma, as there are no diagonals on the circle. Therefore the cylinder is a local as well as global counterexample, and the problem is no longer a case of hidden lemma-incorporation.

Although it would be difficult to model the surprise that a human feels when they attempt to triangulate a circle, the emphasis on vacuously true statements gave us an insight into how to automate this method. We defined the second type of surprise as follows:

- surprisingness (type 2): an entity $m$ which is a global counterexample to a proposition is surprising with respect to one of the conjectures (or lemmas) $C = \forall x(P(x) \rightarrow Q(x))$ in the proof of the proposition, if $\neg P(m)$.

Given a proof-scheme and an entity $m$ which is a global but not local counterexample, our algorithm goes through each lemma $C_i$ in the proof-scheme. If $C_i$ is of the form $\forall x(P(x) \rightarrow Q(x))$, and $\neg P(m)$, then HRL performs two steps: *(i)* it generates the conjecture $C' = \forall x(P(x) \rightarrow Q(x)) \wedge P(m)$ (the entity $m$ is now a counterexample to $C'$), and *(ii)* it returns $C_i$ as the hidden faulty lemma $C$, and $C_i'$ as the explicit lemma.

**Multiple applications of lemma-incorporation**

In the discussion of lemma-incorporation in (Lakatos, 1976), the method is applied to the same conjecture (and proof) at least three times (thus enabling the description of different types of lemma-incorporation). This is like previous methods; one counterexample is found and dealt with and then more counterexamples to the modified conjecture and proof are sought. In HRL, proof-schemes and conjectures are passed around and modified and different students may consider them, or the same student may consider different versions of a proof and conjecture at different times.

### 4.3. Algorithms for lemma-incorporation

We have interpreted Lakatos's method of lemma-incorporation as the series of algorithms shown below. We define $P \rightsquigarrow Q$ to mean that it is *nearly* true that $P \rightarrow Q$, *i.e.*, there are lots of supporting examples and few counterexamples. We implemented these algorithms as a computer program: the teacher in HRL is given a proof-scheme and conjecture $P \rightarrow Q$ by the user, and asks the students to use Lakatos's methods to analyse both proof-scheme and conjecture. Further technical details are given in (Pease, 2007).

1. **Determine which type of lemma-incorporation to perform.** If there are any counterexamples to the global conjecture, then if either: *(i)* these

are also counterexamples to any of the lemmas in the proof, or *(ii)* there is a counterexample to a local lemma, and there is a concept which links the local counterexample to the global counterexample and this concept appears in one of the local lemmas, then perform *global and local lemma-incorporation*. If there is a global counterexample but neither *(i)* nor *(ii)* hold then perform *hidden lemma-incorporation*. Otherwise, if there are counterexamples to any of the lemmas in the proof, then perform *local-only lemma-incorporation*.

2. **Perform local-only lemma-incorporation.** Given a conjecture to which there are no known counterexamples, and a proof tree which contains a faulty lemma, $P \leadsto Q$, to which there is at least one counterexample, then if there is a concept $C$ in the theory which exactly covers the counterexamples (or such a concept can be formed), then make the concept $P \wedge \neg C$, replace the faulty lemma with the conjecture $P \wedge \neg C \rightarrow Q$, and return the improved proof-scheme.

3. **Perform global and local lemma-incorporation.** Given a proof-scheme where there are counterexamples to the global conjecture, $P \leadsto Q$, and these counterexamples are also counterexamples to a lemma in the proof, $R \leadsto S$; form the concept $C$ ‘objects which satisfy the faulty lemma’, by merging the two concepts $R$ and $S$ (this is done by using a production rule to compose $R$ and $S$), modify the global conjecture by making the new conjecture $C \rightarrow Q$, and replace the old global conjecture in the proof-scheme with the modified version.

4. **Perform global-only lemma-incorporation.** Given a proof-scheme where there are counterexamples to the global conjecture, but these counterexamples are not counterexamples to any of the lemmas in the proof, and none of the lemmas have counterexamples which are related to the global counterexamples; let the global conjecture be a near-implication $P \leadsto Q$. Then go through the proof-scheme and take each lemma in turn, testing each to see whether the global counterexample is surprising in the first sense (type 1) with respect to the lemma. If it is, then return this lemma as the hidden faulty lemma and generate another conjecture, to which the global counterexample *is* a counterexample, as the explicit lemma. If not, then go through the proof-scheme and take each lemma in turn, testing each to see whether the global counterexample is surprising in the second sense (type 2) with respect to the lemma. If so, then return this lemma as the hidden faulty lemma, and generate another conjecture, to which the global counterexample *is* a counterexample, as the explicit lemma. If an explicit lemma has been found, then generate an intermediate proof-scheme in which the hidden faulty lemma is replaced by the explicit lemma, perform global and local lemma incorporation on the interme-

diate proof-scheme, and return the result. If no lemma is surprising in either sense, then return the proof-scheme unchanged.

Working in the polyhedra domain, and given the proof tree, conjecture and counterexample as input, HRL has replicated all three of Lakatos's types of lemma-incorporation (Pease, 2007).

## 5. Connections between Haggith and Lakatos

We outline some connections between Haggith's argumentation structures and Lakatos's methods below, which have been suggested by our work in producing a computational reading of Lakatos's theory. In each example, $P$ stands for the proposition "$\forall x(poly(x) \rightarrow euler(x, 2))$", *i.e.*, for all polyhedra, the number of vertices (V) minus the number of edges (E) plus the number of faces (F) is equal to two. Polyhedra for which $V - E + F$ is 2 are called Eulerian. Unless otherwise specified, page references refer to (Lakatos, 1976). In some examples we give alternative propositions for $Q$, which we express as $Q'$. We provide a diagram for each pattern where, again, unmarked arrows represent the justification relation between a proposition and a set of propositions. Given that Lakatos is commonly criticised (for example, Feferman, 1978) for claiming that his methods have general application despite only considering two case studies, we suggest how each argument pattern might describe other mathematical examples.

**Corroboration:** (Figure 7)
$Q$: *All polyhedra in which circuits and bounding circuits coincide, are Eulerian* (p. 114). This is reformulated again to:
$Q'$: *If the circuit spaces and bounding circuit spaces coincide, the number of dimensions of the 0-chain space* minus *the number of dimensions of the 1-chain space* plus *the number of dimensions of the 2-chain space* equals *2* (p. 116).

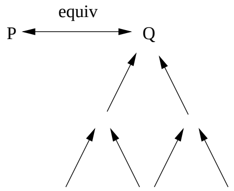


*Figure 7.* Corroboration

This is a reformulation of a problem, where a proposition $P$ is reformulated as proposition $Q$, which is easier to prove. If a convincing proof of $Q$ can be found and it can be shown that $Q$ is equivalent to $P$, then $P$ has been proved. This is a common mathematical technique. Lakatos called this "the problem of translation" and devotes the second chapter of (Lakatos, 1976) to its description, discussion and accompanying problems. The roots of this method lie in Polya's advice: "Could you restate the problem? Could you restate it again?" (Polya, 1945). Analysis of the argument for $Q$ and the premises it contains suggests insights into the original proposition $P$, *i.e.*, ways in which $P$, the concepts in $P$ or the argument for $P$ should be modified.

The trivial case, in which $P$ is equivalent to itself, and multiple proofs are given for the same theorem, is common in mathematics. For instance, Pythagoras's theorem is proved using similar triangles, parallelograms (by Euclid), a trapezoid, similarity, by rearrangement, algebraically, with differential equations and using rational trigonometry. Examples of mathematical statements which are equivalent include: *(i)* Pythagoras's theorem and the parallel postulate; and *(ii)* Zorn's lemma, the well-ordering theorem and the axiom of choice. These examples are particularly interesting since the equivalent statement to $P$, *i.e.*, $Q$, is also a premise in the argument for $P$. That is, Euclid's proof of Pythagoras takes the parallel postulate as a premise, and the proof of the well-ordering theorem uses the axiom of choice. Although this satisfies Haggith's corroboration pattern it would clearly be circular to argue, for instance, that the argument for the well-ordering theorem corroborates the axiom of choice.

There are many examples in mathematics where the relationship between $P$ and $Q$ is not equivalence, but the argument pattern is still relevant: for instance, where $P$ is analogous to, similar to, weaker or more general than, or implies $Q$. Proving special cases of a theorem can be seen as corroborating the theorem: historically we see many examples, such as proving Fermat's Last Theorem for specific values of $n$ (3,5,7), or for classes of number (all regular primes). This can also work the other way around, where a proof of a more general result is easier to find than the proof of the more specific version. For instance, Lagrange's Theorem, that for any finite group G the order (number of elements) of every subgroup H of G divides the order of G, is true under a suitable reformulation even for infinite groups G and H. In this case the proof of the more general statement is simpler than a proof that uses induction over finite structures (some historical details are in Roth, 2001).

**Enlargement** (Figure 8)
$Q$: A polyhedron is a solid whose surface consists of polygonal faces.
$Q'$: A polyhedron is a surface consisting of a system of polygons.

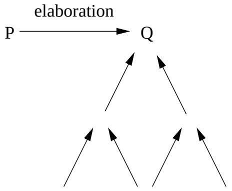


*Figure 8.* Enlargement

Lakatos described both monster-barring, in which a narrow definition of a concept in a conjecture or proposition is proposed in order to exclude a counterexample and thus defend the conjecture, and concept-stretching, in which a wider definition of a concept which covers a counterexample and thus poses problems for the conjecture is proposed. Both methods, in which a concept definition that was previously vague or controversial is made more explicit, can be seen as examples of Haggith's *enlargement* structure. Arguments then given for *why* a particular concept definition should be accepted constitute the argument for $Q$. Lakatos demonstrated the intriguing situation in which we have two enlargement arguments for $P$, with conclusions $Q_1$ and $Q_2$, where $Q_1$ and $Q_2$ disagree. Thus, we might question whether the relationship between $P$ and $Q$ is really an elaboration or not. We show this in Figure 9.

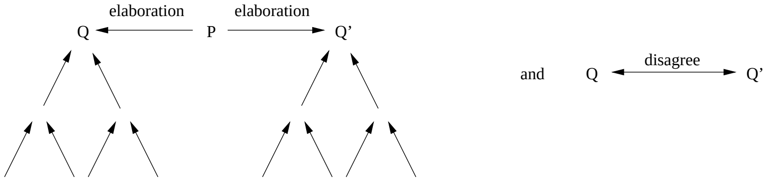


*Figure 9.* Monster-barring and enlargement

In emphasising the role of the argument for $Q$, Haggith stresses the need for *showing*, rather than stating that $Q$ elaborates $P$. This aspect is not always made clear by Lakatos, who did not always clarify how one should select between competing definitions (relying on the literary device of a teacher to state a given definition as fact). This argumentation structure is similar to Walton's argumentation scheme for *argument from verbal classification* (Walton, 2006, pp. 128–132). This is a scheme which takes the form: "$F(a)$, $\forall x(F(x) \rightarrow G(x))$, therefore $G(a)$", where the classifications $F$ and $G$ may be vague or highly subjective.

Other examples of monster-barring or concept-stretching in mathematics include expanding the concept 'number' from natural numbers to include zero, negatives, irrationals, imaginary numbers, transfinite numbers and quaternions; narrowing the concept of 'set' (by limiting the types of sets which can be constructed), Cantor's expansion of our notion of 'size' and changing the definition of 'prime number' from 'a natural number which is only divisible by itself and 1' to 'a natural number with exactly two divisors', thus enabling many theorems about primes, where 1 would have been a counterexample (such as the Fundamental Theorem of Arithmetic), to be neatly stated.

**Consequence:** (Figure 10)
$Q$: A star polyhedron is Eulerian (p. 16).

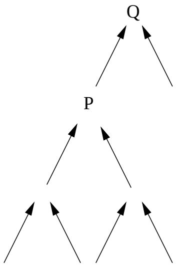


*Figure 10.* Consequence

The proposition $P$, which states that for *all* polyhedra $V - E + F$ is 2, is a premise (the sole premise) in any argument for $Q$ where $Q$ is the proposition that $V - E + F$ is 2 for a particular type of polyhedron. Lakatos used this argumentation structure to great advantage, showing that analysing propositions which are implied by a conjecture is a fruitful way of analysing the conjecture itself. This pattern is similar to corroboration, where there is an implication relation between $P$ and $Q$.

Wiles's proof (with Taylor) for the statement that all rational semistable elliptic curves are modular ($P$) provides another example: this is famous because, as Ribet proved, it implies Fermat's Last Theorem ($Q$). Clearly, proving a given relationship between two mathematical statements is just as important as the argument for one of them. Many examples of lemmas or corollaries in mathematics fit this pattern.

**Rebuttal:** (Figure 11)
$Q$: $\exists x(poly(x) \wedge \neg euler(x, 2))$, or there is an $x$ such that $x$ is a polyhedron and it is *not* the case that the Euler characteristic of $x$ is 2.

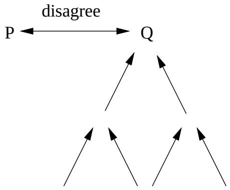


*Figure 11.* Rebuttal

An example is the cylinder (Lakatos, 1976, p. 22), for which $V - E + F$ is 1. This is a case of simple rebuttal, *i.e.* rebuttal without undercutting, which Lakatos addressed in his method of hidden lemma-incorporation. This method demonstrates that it is possible in mathematics to rebut without a known undercutter. This view is discussed (and supported) by Aberdein (2005, p. 298), who argues that the presence of a rebuttal is an *existence proof*, rather than a construction, for an undercutter.

Lakatos also describes how the method of hidden lemma-incorporation was used to fix Cauchy's faulty conjecture that 'the limit of any convergent series of continuous functions is itself continuous' (Lakatos, 1976, app. 1). The counterexample, found by Fourier, is:

$$cosx - \frac{1}{3}cos3x + \frac{1}{5}cos5x - ...$$

which converges to the step function. The most interesting aspect of this example is the timing of various discoveries. Fourier (1808) discovered the above series, and it was *after* this that Cauchy (1821) discovered the conjecture and proof. One solution to this awkward situation was that the limit function was actually continuous, and therefore it was not a counterexample (Fourier held that it was continuous). However, Cauchy had provided a new interpretation of continuity, according to which the limit was *not* continuous (the existence of Fourier's example was considered by some to be evidence that the new interpretation should be rejected). Another possible solution was the argument that the series was not (pointwise) convergent, although this view was not accepted by most mathematicians, including Cauchy, who later proved that it did converge. There was then a long gap until 1847 when Seidel found the hidden assumption of uniform convergence in the proof. Indeed, it was Seidel who invented the method of proofs and refutations. Lakatos thought that the main reason for such a long gap, and the willingness of mathematicians to ignore the contradiction, was a commitment on the part of mathematicians to Euclidean methodology. Deductive argument was considered infallible and therefore there was no place for proof analysis.

A further example can be found in Hilbert's *Grundlagen der Geometrie*:[14]
**Theorem:** For two points A and C there always exists at least one point D on the line AC that lies between A and C.
**Proof:** (paraphrased from Hilbert, 1901, p. 6, as a procedural proof)
*lemma 1:* draw a line AC between the two points
*lemma 2:* mark a point E outside the line AC (axiom (I,3))
*lemma 3:* mark another point F such that F lies on AE and E is a point of the segment AF (axiom (II,2))
*lemma 4:* mark on FC a point G, that does not lie on the segment FC (axiom (II,2) and axiom (II,3))
*lemma 5:* the line EG must then intersect the segment AC at a point D (axiom (II,4))

This proof is accompanied by a diagram containing the hidden assumption that the two points are different (as becomes obvious to humans when they try to draw the line which joins A and C). The counterexample comprises any two points which are identical, *i.e.*, $(a, a)$. With the proof phrased as above,

[14] Since Lakatos described his method in terms of a procedural proof, we paraphrase Hilbert's proof as a procedural proof, from the deductive proof which Hilbert gave. However, it is worth noting that the method of lemma-incorporation also applies to declarative proofs, which we demonstrate with respect to the first step of Hilbert's proof.

$(a, a)$ is a global, but not local counterexample (note again that the *type* of counterexample changes as we step through a proof: in the global theorem the counterexample is a *point*, whereas in lemma one the counterexample is a *line*). This example is interesting, since, in Hilbert's original German edition of the Grundlagen in 1899, reprinted in (Hallett and Majer, 2004), he does not exclude this example: there is nothing in the axioms to say that a line must join two *different* points (the relevant axiom is (I,1), which states that for every two points, A, B there exists a line that contains each of the points. The axiom does not specify that the points must be different), nor that a line which intersects a segment AC must be strictly between the points A and C, etc. However, in later editions, Hilbert has amended this: for instance, at the beginning of (Hilbert, 1901), which is a later, English translation, Hilbert states that in all of the theorems, he assumes that where he says two points, they will be considered to be two distinct points (the relevant omission in his first work is Hallett and Majer, 2004, Chap 5, pp. 437–438). This example was highlighted by Meikle and Fleuriot's (2003) formalisation of (Hilbert, 1901).

Just as mathematical concepts, conjectures and proofs do not emerge fully formed and perfect, neither does axiomatisation of mathematical domains. We believe that Lakatos's methods can be used to describe axiomatisation development in the same way as they describe other mathematical development. Hilbert's axiomatisation of geometry shown here (which itself built on previous work, see Pieri's 1895, 1897–98) is one example. Another is the development of the axioms in set theory, where Frege's comprehension principle is modified to the axiom of subsets, in response to the Burali-Forti, Cantor and Russell paradoxes.

Note that the teacher and students in (Lakatos, 1976) consider the possibility of an entity being a local only, global only, or a both local and global counterexample. They do not consider that a problem entity might be neither global nor local, assuming that such entities are positive examples of the theorem and proof, and therefore support rather than attack it. However, the admission that lemmas in the proof may contain hidden assumptions which mean that an entity satisfies the lemma, albeit in a surprising way, raises the question of whether there could be an entity which satisfies the global conjecture in the same way, thus uncovering a hidden assumption in the global conjecture itself.

**Undermining:** (Figure 12)

$Q$: A polyhedron is a solid whose surface consists of polygonal faces.

$R$: A polyhedron is a surface consisting of a system of polygons.

An example argument of this type is the dialectic over rival definitions in Lakatos's monster-barring method (for further mathematical examples, see

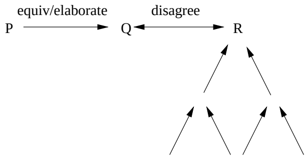


*Figure 12.* Undermining

the section on enlargement.)

**Undercutting:** (Figure 13)
$P_0$: any polyhedron, after having a face removed, can be stretched flat on the blackboard.
$Q$: after removing a face from the hollow cube we cannot stretch it flat on the blackboard.

Alternatively, let
$P_0$: dropping a triangle from a triangulated map always results in either the disappearance of one edge or else of two edges and a vertex.
$Q$: we can drop a triangle from the triangulated map which results from removing a face from the cube and stretching it flat on the blackboard, without it resulting in either the disappearance of one edge or else of two edges and a vertex.

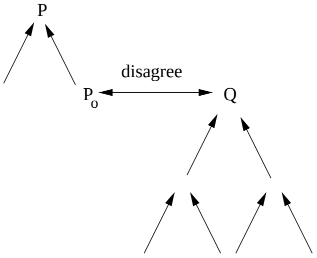


*Figure 13.* Undercutting

This is the earliest argumentation structure to be found in (Lakatos, 1976), where the three steps (or premises) of Cauchy's proof are all questioned on p. 8. The questions are later supported by counterexamples. Note that Lakatos's method of global and local lemma-incorporation is a combination of rebuttal and undercutting, and his local-only lemma-incorporation is just undercutting. Global only, or hidden lemma-incorporation is just rebuttal (as discussed above).

Set theory provides further examples of local-only lemma-incorporation. For instance, Cantor's initial proof that the segment and the square are equivalent sets of points contained the premise that there is a one-to-one onto mapping: $f\ \ \{(x, y)\ \ x \in (0, 1], y \in (0, 1]\} \to (0, 1]$. Cantor identifies such a mapping: let $(x, y)$ be co-ordinates of an arbitrary point in the unit square, where $x = 0.x_1x_2x_3...$, and $y = 0.y_1y_2y_3....$ Then they uniquely determine the point or the unit segment $z = 0.x_1y_1x_2y_2x_3y_3....$ Conversely, every point in the unit segment can be expressed as an infinite decimal. Let $z$ be an arbitrary point in the unit segment, where $z = z_1z_2z_3z_4z_5z_6$... Then this point uniquely determines two co-ordinates $x = z_1z_3z_5$... and $y = z_2z_4z_6$... which determines the points $(x, y)$ of the unit square. Note that Cantor specified that where a number has two decimal expansions, for example 0.2299999999... or 0.230000000, the expansions ending in an infinite set of zeros should be ruled out (so the real number 23/100 is identified as 0.2299999999...). Dedekind (at Cantor's request) checked this proof and found local (but not global) counterexamples. While to each $(x, y)$ there corresponds a single $z$, there exist values of $z$ that arise from no $(x, y)$ in the above procedure. One such counterexample is $z = 0.13050706080$... which yields $x = 0.1$; another counterexample is $z = 0.513020109090$... which yields $y = 0.1$. In these cases neither $x$ nor $y$ is written in admissible decimal form, and if the trailing zeros are replaced by nines then the infinite decimal form does not correspond to the specified number $z$. The proof was then patched so that, instead of considering single digits, groups of digits are considered such that only the last digit of a group differs from 0; so for example $z =$0.1 2 05 0004 2 01 8... would yield $x = 0.10528$... and $y = 0.2000401$... (Burton, 1985, pp. 607–608).

**Target:** (Figure 14)
$Q$: $\exists x(poly(x) \wedge \neg euler(x, 2))$; in particular, the hollow cube has this property, since $V - E + F$ is 4.
$R$: For all polyhedra that have no cavities, $V - E + F$ is 2.
or,
$R'$: $V - E + F = 2 - 2(n-1) + \Sigma_{k=1}^{F} e_k$, for $n$-spheroid — or $n$-tuply connected — polyhedra with $e_k$ edges deleted without reduction in the number of faces (p. 79).

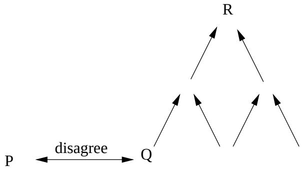


*Figure 14.* Target

Lakatos's exception-barring methods fit this pattern, where a counterexample is found and used, along with other premises, to modify $P$, just producing $R_1$. Another, later example is $R_2$ (p. 79), where many counterexamples have been found and a general repair, $R_2$ has been found which explains all of the positive and negative examples. Lakatos called this the problem of content, which deals with the problem that with every repair, the domain of application of the conjecture (originally all polyhedra, and then all polyhedra of an increasingly narrow type) has narrowed to the stage where the conjecture is no longer very interesting. Thus, mathematicians were "striving for truth at the expense of content" (p. 66). In this case, Lakatos suggested (again building on work by Polya, 1945) distinguishing an initial problem in which a question is raised, from an initial conjecture which poses a first answer to the initial problem. If this answer is refined almost to the point of a tautology (approaching the 'conjecture' that any Eulerian polyhedron is Eulerian) then it may be preferable to return to the initial problem and pose another answer.

Historical conjectures about perfect numbers provide an example in number theory. For instance, using the (scientific) inductive argument that because the first four perfect numbers, 6, 28, 496 and 8128 each contain $n$ digits, it was conjectured that the $n$th perfect number $P_n$ contains exactly $n$ digits $(P)$. The fifth perfect number, 33,550,336, provides a counterexample $(Q)$, the discovery of which led to the new conjecture that perfect numbers end alternately in 6 and 8 $(R)$. Again, the argument pattern repeated itself with the discovery of the sixth perfect number, 8,589,869,056, which is a counterexample to $R$. This led to further statements and arguments, including the theorem that the last digit of any even perfect number must be 6 or 8 and the (open) conjecture that all perfect numbers are even.

**Counter-consequence:** (Figure 15)
$Q$: $\neg\exists x(poly(x) \wedge \neg euler(x, 2))$.
$R$: $\exists x(poly(x) \wedge \neg euler(x, 2))$; in particular, the twin tetrahedra polyhedron has this property, since it has an Euler characteristic of 3 (p. 15).
$S$: for any polyhedron that has no 'multiple structure', $V - E + F = 2$.

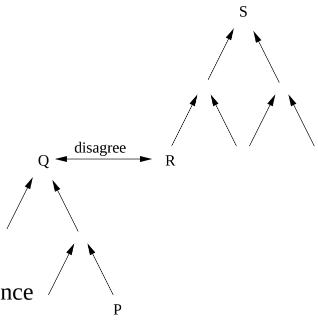


*Figure 15.* Counter-consequence

This is very similar to the *target* structure described above. Again, Lakatos's exception-barring methods present an example of this type of structure. The perfect number examples described above also fit this pattern. Other examples of monster-adjusting seem rare in mathematics. However, we can see monster-adjusting as a type of monster-barring, where the concept in question may be the right hand concept in an implication or equivalence conjecture, rather than the domain. Let us formalise monster-barring as follows: from conjecture $\forall x(P(x) \rightarrow Q(x))$, and (known) counterexample $m$ such that $P(m)$ and $\neg Q(m)$, (re)define either $P$ or $Q$ so that for the $m$ in question, either $\neg P(m)$ or $Q(m)$ is true. Monster-adjusting can now be seen as a case of this formalisation, where the concept under debate is $Q$ rather than $P$.

It is interesting that the structures which Haggith has identified can be used to express most of Lakatos's methods. A method which has so far been neglected is monster-adjusting. As with monster-barring, this method also exploits ambiguity in concepts, but reinterprets an object in such a way that it is no longer a counterexample. The example in (Lakatos, 1976) concerns the star polyhedron. This entity is raised as a counterexample since, it is claimed, it has 12 faces, 12 vertices and 30 edges (where a single face is seen as a star polygon), and thus $V - E + F$ is $-6$. This is contested, and it is argued that it has 60 faces, 32 vertices and 90 edges (where a single face is seen as a triangle), and thus $V - E + F$ is 2. The argument then turns to the definition of 'face'. A new structure which corresponds to this method might look like the one shown in Figure 16, where:
$Q$: $\exists x(poly(x) \wedge \neg euler(x, 2))$; in particular, the star polyhedron has this property, since $V - E + F$ is $-6$ (p. 16).
$R$: The star polyhedron has 12 (pentagonal) faces. (Other premises in the argument for $Q$ are that the star polyhedron has 12 vertices and 30 edges.)
$S$: The star polyhedron has 60 (triangular) faces (p. 31).

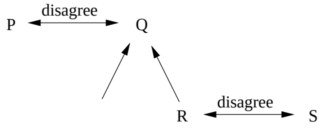


*Figure 16.* A new argumentation structure for monster-adjusting

## 5.1. Mutually beneficial associations between Lakatos and Haggith

### Using Haggith's work to extend Lakatos's theory

One shortcoming of Lakatos's representation of an initial proof, as a list of lemmas, or conjectures, is that it fails to show how the conjectures in the proof fit together. In both of his case studies, the proof is represented by Lakatos as a series of local conjectures which, taken together, imply the global conjecture. In Haggith's notation, this would be simply written as $A = \{C \Leftarrow \{P1, P2, P3\}\}$. This prevents the expression of any dependencies between the conjectures and thus would be an extreme oversimplification of most mathematical proofs, even at an initial stage. Using Haggith's notation makes explicit, rather than hides, the structure of a proof. Seeing how a proof fits together may make it easier to identify flaws and their solution.

### Using Lakatos's work to extend Haggith's theory

Using Lakatos's case studies, we have shown that Haggith's argumentation structures, which were inspired by the need to represent different perspectives in natural resource management, can be usefully applied to mathematical examples. This is a new domain for Haggith, and thus supports her claim that the structures are of general use.

Combining Lakatos's conjecture-based and Haggith's proposition-based representations has the advantage of highlighting weak areas in a proof. This might be in the relationships between sets of conjectures, such as justification or elaboration, or in the claims asserted by the conjectures. Propositions are no longer black boxes, thus enabling new areas for flaws to be found and repaired. Lakatos's methods, for example his "problem of content", also consider the interestingness of a proposition.

Lakatos's methods suggest new structures for Haggith. Although she made no claim to have identified all structures, adding new examples to the catalogue is a valuable contribution to Haggith's work.

## 6. Conclusion

Many argumentation theorists have assumed, either implicitly or explicitly, that mathematics is about formal reasoning and is therefore not a suitable domain for argumentation. We hope that the ideas in this paper, in addition to other work such as (Aberdein, 2005), will show that this assumption is not justified and that mathematics is a rich and fertile domain for informal reasoning techniques. In accordance with the computational philosophy paradigm, we believe that implementing Lakatos's theory has provided a new perspective on it. One insight is discovering the utility of argumentation structures for proof-schemes, which has allowed us to represent them, as well as suggesting ways in which Lakatos's methods for repairing faulty conjectures and proofs can be implemented within our model. Thus, the process of producing computational models of theories can both suggest and exploit new connections.

Just as Lakatos suggested that embedding a conjecture in a different body of knowledge can lead to further insights into the conjecture, so we can see the value of embedding philosophy of mathematics in argumentation theory and vice versa. These two normally unconnected domains have much to offer each other. The computational modelling domain adds a further dimension of interest and will play an essential role in making these intriguing connections richly detailed and explicit.

## Authors' Vitae

*Dr. Alison Pease*
is a Research Associate on the project "A Cognitive Model of Axiom Formulation and Reformulation with applications to AI and Software Engineering" at the University of Edinburgh. She obtained her PhD in 2007 after completing an MA in Mathematics/Philosophy from the University of Aberdeen and an MSc in Artificial Intelligence from the University of Edinburgh. Her PhD thesis was entitled "A Computational Model of Lakatos-style Reasoning" and described her implementation and evaluation of Lakatos's theory of mathematical progress. Previous to obtaining her PhD Alison worked as a mathematics teacher for four years, teaching levels from Special Needs to A level students. She holds a PGCE specialising in teaching mathematics.

*Dr. Alan Smaill*
has worked in the area of Automated Reasoning since 1986. He holds a D.Phil. from the University of Oxford in Mathematical Logic, and is currently a lecturer in the Division of Informatics in the University of Edinburgh. His research work centres around reasoning in higher-order and constructive logics, with applications in program construction and automated software engineering, and he has published widely in this area. He is currently a Principal Investigator on an EPSRC grant on the mechanisation of first-order temporal logic via embedding into a higher-order representation and also on "A Cognitive Model of Axiom Formulation and Reformulation with applications to AI and Software Engineering".

*Dr. Simon Colton*
is a senior lecturer in the Department of Computing at Imperial College London. He runs the Combined Reasoning Group (`www.doc.ic.ac.uk/crg`), which investigates ways in which to integrate Artificial Intelligence systems so that the whole is more than a sum of the parts. The group test their techniques in applications which usually require an aspect of creative behaviour in the software, which include projects in mathematical discovery, graphic design, bioinformatics, visual arts and video game design. Dr. Colton's research has been recognised at both a national and international level, with the BCS/CPHC distinguished dissertation award in 2001 and the AAAI best paper award in 2000.

*Dr. John Lee*
is Deputy Director at the Human Communication Research Centre, which links Informatics, Linguistics and Psychology at Edinburgh, and with the Universities of Glasgow and Durham. He is also Co-ordinator of the Edinburgh-Stanford Link, which specialises in speech and language technology. His

research interests include multimodal dialogue, graphics in reasoning and learning, computing and cognition in design. In particular, he has investigated the paradigm of “Vicarious Learning”, the relationship between language and other modalities, information technology in support of design practice, using automated synthesis of web sites to communicate about accident reports. He holds a PhD in Philosophy and Cognitive Science from the University of Edinburgh.

## Acknowledgements

We would like to thank our three reviewers who all gave comprehensive, constructive and thought-provoking reviews. Andrew Aberdein also made very helpful comments. This work was partly funded by EPSRC grant EP/F035594/1.

## References

Aberdein, A.: 2005, ‘The Uses of Argument in Mathematics’. *Argumentation* **19**(3), 287–301.

Aberdein, A.: 2006, ‘Managing Informal Mathematical Knowledge: Techniques from Informal Logic’. In: J. M. Borwein and W. M. Farmer (eds.): *MKM 2006*, LNAI 4108. Berlin: Springer-Verlag, pp. 208–221.

Alcolea Banegas, J.: 1998, ‘L’Argumentació en Matemàtiques’. In: E. Casaban i Moya (ed.): *XIIè Congrés Valencià de Filosofia*. Valencià: pp. 135–147.

Burton, D.: 1985, *The History of Mathematics*. Boston: Allyn and Bacon.

Cauchy, A. L.: 1813, ‘Recherches sur les Polyèdres’. *Journal de l’École Polytechnique* **9**, 68–86.

Cauchy, A. L.: 1821, *Cours d’Analyse de l’École Royale Polytechnique*. Paris: de Bure.

Colton, S.: 2002, *Automated Theory Formation in Pure Mathematics*. New York: Springer-Verlag.

Corfield, D.: 1997, ‘Assaying Lakatos’s Philosophy of Mathematics’. *Studies in History and Philosophy of Science* **28**(1), 99–121.

Crawshay-Williams, R.: 1957, *Methods of Criteria of Reasoning: An Inquiry into the Structure of Controversy*. London: Routledge and Kegan Paul.

Crelle, A. L.: 1826-1827, *Lehrbuch der Elemente der Geometrie*, Vol. 1,2. Berlin: Reimer.

Elvang-Gøransson, M., P. Krause, and J. Fox: 1993, ‘Dialectical reasoning with inconsistent information’. In: *Proceedings of the 9th Conference on Uncertainty in AI*. San Mateo, CA: Morgan Kaufmann, pp. 114–121.

Ernest, P.: 1997, ‘The Legacy of Lakatos: Reconceptualising the Philosophy of Mathematics’. *Philosophia Mathematica* **5**(3), 116–134.

Feferman, S.: 1978, ‘The Logic of Mathematical Discovery vs. the Logical Structure of Mathematics’. In: P. D. Asquith and I. Hacking (eds.): *Proceedings of the 1978 Biennial Meeting of the Philosophy of Science Association*, Vol. 2. East Lansing, MI: Philosophy of Science Association, pp. 309–327.

Fourier, J.: 1808, ‘Mémoire sur la Propagation de la Chaleur dans les Corps Solides (Extrait)’. *Nouveau Bulletin des Sciences, par la Société Philomathique de Paris* **1**, 112–16.

Goguen, J.: 1999, ‘An introduction to algebraic semiotics, with application to user interface design’. In: C. L. Nehaniv (ed.): *Computation for Metaphors, Analogy, and Agents*, LNAI 1562. Berlin: Springer-Verlag, pp. 242–291.

Haggith, M.: 1996, ‘A meta-level argumentation framework for representing and reasoning about disagreement’. Ph.D. thesis, Dept. of Artificial Intelligence, University of Edinburgh.

Hallett, M. and U. Majer (eds.): 2004, *David Hilbert’s Lectures on the Foundations of Geometry: 1891-1902*. Berlin: Springer-Verlag.

Hardy, G. H.: 1928, ‘Mathematical Proof’. *Mind* **38**, 11–25.

Hilbert, D.: 1901, *The Foundations of Geometry*. Open Court. English translation by E. J. Townsend.

Jonquières, E.: 1890, ‘Note sur un Point Fondamental de la Théorie des Polyèdres’. *Comptes Rendus des Séances de l’Académie des Sciences* **110**, 110–115.

Lakatos, I.: 1976, *Proofs and Refutations*. Cambridge: Cambridge University Press.

Larvor, B.: 1998, *Lakatos: An Introduction*. London: Routledge.

Matthiessen, L.: 1863, ‘Über die Scheinbaren Einschränkungen des Euler’schen Satzes von den Polyedern’. *Zeitschrift für Mathematik und Physik* **8**, 1449–450.

Meikle, L. and J. Fleuriot: 2003, ‘Formalizing Hilbert’s Grundlagen in Isabelle/Isar’. In: D. Basin and B. Wolff (eds.): *Proceedings of the 16th International Conference on Theorem Proving in Higher Order Logics*, LNCS 2758. Berlin: Springer-Verlag, pp. 319–334.

Naess, A.: 1953, *Interpretation and Preciseness: A Contribution to the Theory of Communication*. Oslo: Skrifter utgitt ar der norske videnskaps academie.

Naess, A.: 1966, *Communication and Argument: Elements of Applied Semantics*. London: Allen and Unwin. Translation of *En del Elementaere Logiske Emner*. Universitetsforlaget, Oslo, 1947.

Pease, A.: 2007, ‘A Computational Model of Lakatos-style Reasoning’. Ph.D. thesis, School of Informatics, University of Edinburgh: `http://hdl.handle.net/1842/2113`.

Pease, A., S. Colton, A. Smaill, and J. Lee: 2004, ‘A Model of Lakatos’s Philosophy of Mathematics’. *Proceedings of Computing and Philosophy (ECAP)*.

Pedemonte, B.: 2000, ‘Some cognitive aspects of the relationship between argumentation and proof in mathematics’. In: M. van den Heuvel-Panhuizen (ed.): *25th Conference of the International Group for the Psychology of Mathematics Education*. Utrecht.

Pollock, J.: 1970, ‘The structure of epistemic justification’. In: N. Rescher (ed.): *Studies in the Theory of Knowledge*, American Philosophical Quarterly Monograph Series 4. Oxford: Blackwell, pp. 62–78.

Pollock, J.: 1995, *Cognitive Carpentry*. Cambridge, MA.: The MIT press.

Polya, G.: 1945, *How to solve it*. Princeton, NJ: Princeton University Press.

Polya, G.: 1954, *Mathematics and Plausible Reasoning: Vol. 1, Induction and Analogy in Mathematics*. Princeton, NJ: Princeton University Press.

Popper, K. R.: 1959, *The Logic of Scientific Discovery*. New York: Basic Books.

Reed, C. and G. Rowe: 2005, ‘Translating Toulmin Diagrams: Theory Neutrality in Argument Representation’. *Argumentation* **19**(3), 267–286.

Roth, R. L.: 2001, ‘A History of Lagrange’s Theorem on Groups’. *Mathematics Magazine* **74**(1), 99–108.

Russell, B.: 1971, *Logic and Knowledge: Essays 1901-1950*. London: George Allen and Unwin.

Sartor, G.: 1993, ‘A Simple Computational Model for Nonmonotonic and Adversarial Legal Reasoning’. In: *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*. Amsterdam: ACM.

Toulmin, S.: 1958, *The Uses Of Argument*. Cambridge: Cambridge University Press.

Toulmin, S., R. Rieke, and A. Janik: 1979, *An Introduction to Reasoning*. London: Macmillan.

Walton, D.: 2006, *Fundamentals of Critical Argumentation*. Cambridge: Cambridge University Press.

Wilder, R. L.: 1944, 'The Nature of Mathematical Proof'. *The American Mathematical Monthly* **51**(6), 309–323.

*Address for Offprints:* Alison Pease,
Informatics Forum, University of Edinburgh,
10 Crichton Street, Edinburgh, EH8 9AB, UK
Tel: +44 (0)131 650 2725
Fax: +44 (0)131 650 6899
Email: A.Pease@.ed.ac.uk