Logic and Logical Philosophy
Volume 20 (2011), 7–57
DOI: 10.12775/LLP.2011.002

Alison Pease
Andrew Aberdein

FIVE THEORIES OF REASONING: Interconnections and applications to mathematics

Abstract. The last century has seen many disciplines place a greater priority on understanding how people reason in a particular domain, and several illuminating theories of informal logic and argumentation have been developed. Perhaps owing to their diverse backgrounds, there are several connections and overlapping ideas between the theories, which appear to have been overlooked. We focus on Peirce's development of abductive reasoning [39], Toulmin's argumentation layout [52], Lakatos's theory of reasoning in mathematics [23], Pollock's notions of counterexample [44], and argumentation schemes constructed by Walton *et al.* [54], and explore some connections between, as well as within, the theories. For instance, we investigate Peirce's abduction to deal with surprising situations in mathematics, represent Pollock's examples in terms of Toulmin's layout, discuss connections between Toulmin's layout and Walton's argumentation schemes, and suggest new argumentation schemes to cover the sort of reasoning that Lakatos describes, in which arguments may be accepted as faulty, but revised, rather than being accepted or rejected. We also consider how such theories may apply to reasoning in mathematics: in particular, we aim to build on ideas such as Dove's [13], which help to show ways in which the work of Lakatos fits into the informal reasoning community.

Keywords: informal reasoning, mathematics, Lakatos, argumentation.

1. Introduction. Informal reasoning

1.1. Lessons from Norbury. In one of the few cases in which Sherlock Holmes was mistaken, "The adventure of the yellow face" [11, pp. 28–46],

Special Issue: *Logic in Cognitive Science*. Guest Editors: J. Malinowski and R. Palczewski

ISSN: 1425-3305

Holmes's client, Grant Munro, asks the detective for help in deciphering his wife's recent odd behaviour. Munro's wife Effie was previously married and lost her first husband and child to yellow fever, but now lives in Norbury and is seemingly happy with her life and new husband. Lately, however, she has requested a large sum of money, without elaborating on her reasons, undertaken mysterious outings in the middle of the night, been spotted secretly visiting the occupants of a neighbouring, only very recently occupied cottage and generally begun to behave in an extremely guilty fashion. Furthermore, Munro has observed, in strange circumstances, what appears to be the face of a man watching him through a window of the neighbouring cottage (though it was viewed from afar, and thus he cannot be sure that it was a man). From this information, Holmes concludes, provisionally, that the mysterious figure is Effie's first husband, who in fact did not die, and is now blackmailing her in some way. The truth is revealed later the same day in the form of Effie's daughter, and it transpires that she was the mysterious figure at the window: the first husband *did* in fact die, but the daughter, whom Effie was trying to keep secret, survived. These circumstances lead to a happy ending, with Munro embracing the child and accepting her as one of his own. Holmes's flawed reasoning prompts one of his most humble quotes:

> Watson, if it should ever strike you that I am getting a little overconfident in my powers, or giving less pains to a case than it deserves, kindly whisper 'Norbury' in my ear, and I shall be infinitely obliged to you. [11, p. 46]

Various aspects of the type of reasoning which Holmes follows have been elucidated by informal logicians or argumentation theorists. They hold that it is useful to describe the kind of reasoning that people perform (as contrasted with normative classical logic). In particular, the last hundred or so years have seen a big surge in the development of informal logic. For instance, Peirce [39] developed the notion of abduction, reasoning from effect to cause, which underpins all Holmesean arguments. Toulmin [52] picked apart argumentation structures and identified different types of statement which in some way support a claim (the mysterious figure is Effie's first husband), such as a specific datum (Effie has secretly visited the occupants of the cottage), a general statement, or warrant (if one's wife secretly goes out then she is hiding something), and a qualifier on the claim (Holmes believed it provisionally, with a view to testing his hypothesis). Lakatos [23] describes the type of reasoning that mathematicians employ, in particular characterising different reactions to counterexamples (in the story above Holmes was quick to surrender his hypothesis when faced with the daughter, but in other situations, one may continue the train of thought — perhaps the daughter *and* the first husband are hiding in the house?). Pollock [44] identified different kinds of counterexample, defeaters that deny a conclusion, and defeaters that attack the reason for believing the claim, but not the claim itself. This type of reasoning is embedded in the story above: if the claim is that there is a man in the cottage, based on the reason that Munro saw a man in the window, then the fact that he was far away at the time, and thus could not see very well, attacks the reason for believing his claim (although there may still be a man in the cottage). Walton and colleagues [55, 54] present a catalogue of argumentation schemes and corresponding critical questions, including *Argument from Evidence to a Hypothesis* [54, pp. 331f.] (see §7.5.1) which is perhaps closest to capturing Holmes's reasoning, but also many other schemes such as *Argument from Expert Opinion* [54, pp. 312 f.] or *Ethotic Argument* [54, p. 336], describing how we may place more weight on Holmes's conclusion than on that of others, since he is so rarely mistaken.[1]

1.2. The blind men and the elephant. It is probably impossible to ever give a comprehensive theory of informal reasoning. In this regard we are like blind men attempting to understand an elephant by each touching a different part.[2] The challenge is made more acute by its strong interdisciplinary nature: many theorists of argumentation bring to the subject a background in a particular domain, often law or philosophy, but also science, medicine, mathematics, ethics, AI, etc. This background is often reflected in the type of argumentation discussed (despite some claims of generality). Our aim in this paper is to attempt to overlay some of the most important descriptions of the elephant, exploring similarities both

[1] The publication dates by the authors we discuss do not always reflect when they were written. Roughly speaking, and to the nearest decade, Peirce developed his ideas on abduction from the 1860's onwards, Lakatos worked on heuristics for mathematical discovery during his Ph.D. in the late 1950's, Toulmin developed (and published) his work on argument in the late 1950's, Pollock started work on defeasible logic and epistemic justification in the late 1960's, and Walton embarked on his program of argumentation schemes in the 1990's

[2] For an unusually detailed application of the fable to mathematics in general, see [35, p. 38f.].

between different descriptions and within a single description. In Part I we summarise some salient theories, and in Part II we consider some connections between them. (Readers who are familiar with the work of Peirce, Toulmin, Pollock, Lakatos and Walton may safely proceed straight to Part II.) We consider the theorists mentioned above who, despite being amongst the best-known writers on informal reasoning, make surprisingly few references to each other's work (excepting perhaps Walton, who brings a contemporary flavour to the paper). The interdisciplinary nature provides a likely explanation of this oversight: different motivations behind the development of their theories may have led to relevant work being unread and some connections going unseen (for example, in law the focus of argumentation might be on particular cases, while in mathematics it might be a general rule that is tested).

Because of the importance of background motivation in guiding description, we must be honest about our own motivation. One of our long-term goals is to show how theories of informal reasoning can be applied to mathematical reasoning, an oft-neglected domain. This goal contrasts claims, often by the theorists themselves, that mathematical reasoning enjoys a special status and is exempt from such theories. We take the opportunity to further our aim in this paper, by suggesting how such theories may apply to mathematics and using examples from mathematics, where possible. Because of this focus we emphasise Lakatos's work in particular. Possibly because he published in the philosophy of mathematics, none of the other authors we consider ever referenced his work, so far as we have been able to determine. Thus we are keen to point out the relevance of Lakatos's work to the informal reasoning community and to firmly embed his work within the field.[3]

1.3. A note on terminology. Argumentation theorists sometimes write as though an argument has been pre-constructed and is now in a presentational format designed for communication (perhaps originating with Aristotle's distinction between rhetorical syllogisms which are aimed at persuasion and scientific syllogisms which are aimed at demonstration, both a form of communication). This contrasts with work on informal logic or logics of discovery, in which historical case studies are analysed in order to glean general heuristics for discovery or reasoning at an individual level. In this paper we do not intend any such distinction, seeing an

[3] A goal that others share: see, for example, [13].

argument as a possibly private chain of thought, and thus argumentation theory as a subfield of informal logic.

Part I. Some salient theories

2. Peirce's theory of abduction

Peirce was educated as a chemist and worked for thirty years as a scientist. In addition, he is also considered to be a philosopher, logician and mathematician.

2.1. Peirce and argumentation Peirce dissected argument into three "utterly irreducible" [39, 2.146][4] forms of reasoning: deduction, induction and abduction. He saw these three forms as the only elementary modes of reasoning there are [39, 8.209], and the building blocks for all other arguments: for instance, he saw analogy as consisting of two stages involving (*i*) induction and then deduction, and (*ii*) abduction and then deduction [39, 2.513]. He distinguished these forms in the following way [39, 2.623]:

2.1.1. Deduction
Rule All the beans from this bag are white.
Case These beans are from this bag.
∴ **Result** These beans are white.

2.1.2. Induction
Case These beans are [randomly selected] from this bag.
Result These beans are white.
∴ **Rule** All the beans from this bag are white.

2.1.3. Hypothesis [Abduction]
Rule All the beans from this bag are white.
Result These beans [oddly] are white.
∴ **Case** These beans are from this bag.

The three forms can also be described in terms of the three propositions below, where deduction uses *modus ponens* to reason from rule (3) and specific case (1), to (2).

[4] Following convention, our references to Peirce's *Collected Papers* cite volume and paragraph number, rather than page number.

$$Pm \tag{1}$$
$$Qm \tag{2}$$
$$\forall x(Px \to Qx) \tag{3}$$

It is explicative (analytic) and valid (the conclusion necessarily follows from the premisses). Induction reasons from many specific cases of (1) and (2), to general rule (3). (It may be an equivalence relation or the implication may be reversed.) It is ampliative (synthetic) and invalid (the conclusion does not necessarily follow from the premisses). *Mathematical* induction, of course, is deductive reasoning. For a given property P and natural number n, a mathematician would use the first rule below and show that the second holds, in order to conclude, via *modus ponens*, the third:

$$((\forall n(Pn \to P(n+1))) \wedge P0) \to \forall x Px$$
$$(\forall n(Pn \to P(n+1))) \wedge P0$$
$$\forall x Px$$

Abduction reasons from specific case (2) and rule (3) to a possible explanation for (2), *e.g.*, (1). It is ampliative (synthetic) and invalid. The conclusion amplifies rather than explicates what is stated in the premisses.

Peirce thought [39, 8.384] that there are two important characteristics in each type of reasoning — *security* (the level of confidence we have in inferences), and *uberty* (the value in productiveness). Roughly speaking, these respectively decrease and increase from deduction to induction to abduction.

2.2. A focus on abduction. For the purposes of this paper we focus on abduction. This was regarded by Peirce as the cornerstone of all scientific discovery: he argued that "Every single item of scientific theory which stands established today has been due to Abduction." [39, 8.172], and "All the ideas of science come to it by way of Abduction." [39, 5.145] Peirce's theory of abduction (also called Hypothesis, Retroduction and Hypothetic Inference) evolved during his life, from about 1860 to 1911 and has since been developed, for example by Hanson [20]. There is no single accepted account. Peirce [39, 5.188–89] describes its form as:

> The surprising fact C is observed;
> But if A were true, C would be a matter of course;
> Hence, there is reason to suspect that A is true.

Abduction consists of two tasks (these tasks are somewhat conflated in the literature, with abduction being defined at different times as one or the other):

- generation of different hypotheses
- selection of best hypothesis (to start testing)

Note that the selection in abduction only determines which hypothesis we take *on probation*, or "set down upon our docket to be tried" [39, 5.602]. This is contrasted with eventual acceptance of a 'hypothesis'—normally called something stronger such as 'scientific law'—which we hold to be true (at least for some time).

Peirce claimed that reasons for carrying out both of these tasks are analogous and determined by what we intend to do with the hypothesis. He argues that the central problem of abduction therefore comes down to understanding the criteria for selection of the best hypothesis. These criteria are:

1. it must explain the surprising fact
2. it must be subject to experimental testing
3. it must be economical; *i.e.*, it must be worth our time to investigate. We should consider:
   (a) the cost of verifying or falsifying the hypothesis (should be low);
   (b) the intrinsic value in the hypothesis (should be high)—where value is defined as (*i*) its *simplicity*—follow the principle of Ockham's razor (all else being equal, prefer the simplest explanation); and (*ii*) the *likelihood* of it being true (estimated by previous experience).
   (c) the effect of the hypothesis on other projects

These are in order of priority. The first two, which Peirce considered to be either satisfied or not, must be fulfilled before the third, which is on a sliding scale, is considered. Since it is likely that we'll have to test (by induction) many hypotheses, it is important that we select those which are most economical.

2.3. The role of surprise. Peirce attributes an important role to surprise: "it is by surprise that experience teaches all she deigns to teach us" [39, 5.51]. However, although he assumes that the initial fact to be explained is surprising, this is not a necessary criterion in more recent writing on abduction. For instance, diagnosis (reasoning from symptoms to possible causes) is seen as an example of abductive reasoning, yet it is not necessarily *surprising* that someone has certain symptoms. There is ongoing debate in the creativity world on the role of surprise and its importance for discovery (see, for example [6, 32]).

2.4. Three stages of enquiry In Peirce's later writings the three types of reasoning become three stages of enquiry. Firstly abduction is used to explain facts; secondly deduction is used to draw out testable consequences of the explanations (make predictions); and finally induction is used to test these consequences (compare predictions with observed behaviour), thereby providing support for the explanation by failing to falsify it, or disproving it by falsification.

3. Toulmin's argumentation layout

Toulmin was a philosopher, author and educator, with a background in moral reasoning: he analysed practical arguments which evaluated moral issues. In Toulmin's well-known model of argumentation [52], written as a critique of formal logic, he argued that practical arguments focus on justification rather than inference. He 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 identified different types of statement which in some way *support* a claim C (the conclusion of the argument): data D (facts we appeal to as the foundation of the claim, or minor premise), warrant W (the statement authorising the move from the data to the claim, or major premise), backing B (further reason to believe the warrant), and a qualifier Q (such as "probably", "certainly" or "necessarily", which expresses the force of the claim). There is only one type of statement in Toulmin's layout which *opposes* a claim: a rebuttal R, which Toulmin defines as "the exceptional conditions which might be capable of defeating or rebutting the warranted conclusion" [52, p. 101]. (Reed and Rowe argue that 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 [47, pp. 15–19].) We show Toulmin's general layout, from [52, p. 103], in Figure 1 (initially the backing and warrant were on the same side as the rebuttal, at some point during Toulmin's development, this flipped to the orientation we present here). Briefly speaking:

- B is reason to believe W;
- W and D together are reasons to believe C;
- R tells you when belief in C is *not* supported; and
- Q is *how much* to believe C.

This may be illustrated by one of Toulmin's best known examples: Given that *Harry was born in Bermuda*^{D}, we can *presumably*^{Q} claim that *he is British*^{C}, since *anyone born in Bermuda will generally be British*^{W} (on account of *various statutes ...*^{B}), unless *his parents were aliens, say*^{R} (derived from [52, p. 104]).

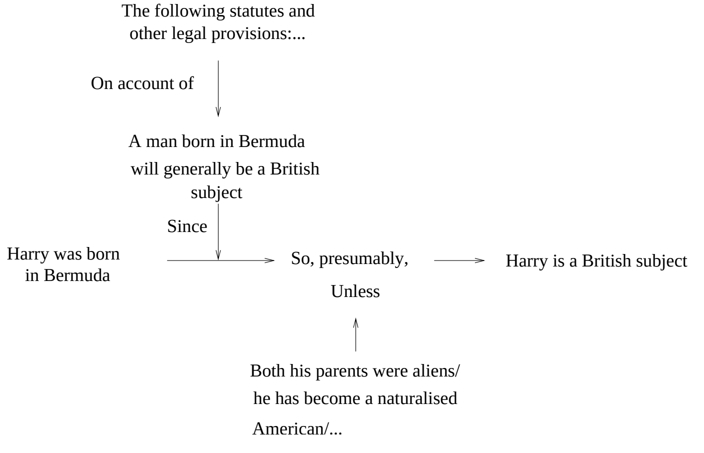

Figure 1. Toulmin's layout.

4. Pollock's two defeaters

Pollock was a philosopher who contributed to epistemology, philosophical logic, cognitive science and AI. He pointed out [44] that within epistemic reasoning, some reasons are *conclusive reasons*, i.e. they logically entail their conclusions. But there are also *non-conclusive reasons* within epistemic reasoning that support their conclusions only defeasibly. These are *prima facie* reasons. There are two types of considerations that defeat *prima facie* reasons: defeaters that deny a conclusion P, and defeaters that attack the reason for believing P, but not P itself. Pollock's general definition of a defeater is "If P is a reason for S to believe Q, R is a defeater for this reason if and only if R is logically consistent with P and $(P \wedge R)$ is not a reason for S to believe Q". He then specifies rebutting and undercutting defeaters as follows:

> If P is a *prima facie* reason for S to believe Q, R is a *rebutting* defeater for this reason if and only if R is a defeater (for P as a reason for S to believe Q) and R is a reason for S to believe not-Q;
>
> If P is a *prima facie* reason for S to believe Q, R is an *undercutting* defeater for this reason if and only if R is a defeater (for P as a reason for S to believe Q) and R is a reason to deny that P would not be true unless Q were true [43, pp. 38 f.].

Hence a *rebutting defeater* is a reason for denying a claim which is supported by a prima facie reason; and an *undercutting defeater* attacks the connection between a prima facie reason and the conclusion, rather than attacking the conclusion directly[5]. This latter is interesting since it does not offer any support for statements which refute the conclusion either.

His example argument is as follows: "suppose x looks red to me, but I know that x is illuminated by red lights and red lights can make objects look red when they are not. Knowing this defeats the prima facie reason, but it is not a reason for thinking that x is *not* red" [44, p. 41].

Pollock claims [44, p. 41] that rebutting and undercutting defeaters are the only two kinds of defeater necessary for describing the full complexity of defeasible reasoning.

5. Lakatos's heuristics for discovery

Lakatos was a philosopher of mathematics and science. He held a fallibilist picture of mathematics, and 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. He demonstrated his argument by presenting case studies 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

[5] Pollock claims [44, p. 41] to have been the first, in [42], to explicitly point out defeaters other than a rebuttal. However, Pollock does not seem to have been aware of Lakatos's work: [23] was initially published between May 1963 and August 1964 as a series of articles for the *British Journal for the Philosophy of Science* [25, 26, 27, 28], appearing six years before [42]. Since Lakatos's work was in the philosophy of mathematics, there is no reason why Pollock should have been aware of it.

functions is itself continuous. In [23] Lakatos presented a rational reconstruction of the history of ideas in the philosophy of mathematics as well as these two mathematical conjectures: this traces psychologism, intuitionism, rationalism, historicism, pragmatism, dogmatism, Kant's idea of infallible mathematics, refutationism, inductivism and deductivism. He presented his work as a classroom discussion between (very advanced) students: we follow his convention of naming the students with letters from the Greek alphabet. We outline Lakatos's methods below.

The method of **surrender** consists of abandoning a conjecture in the light of a counter-example.

Piecemeal exclusion deals with exceptions by excluding a whole class of counterexamples. This is done by generalising from a counterexample to a class of counterexamples which have certain properties. For instance, the students generalise from the hollow cube to *polyhedra with cavities*, and then modify Euler's conjecture to 'for any polyhedron without cavities, $V - E + F = 2$'.

Strategic withdrawal uses positive examples of a conjecture and generalises from these to a class of object, and then limits the domain of the conjecture to this class. For instance, the students generalise from regular polyhedra to *convex polyhedra*, and then modify Euler's conjecture to 'for any convex polyhedron, $V - E + F = 2$'.

Monster-barring/monster-adjusting is a way of excluding an unwanted counterexample. This method starts with the argument that a 'counterexample' can be ignored because it is *not* a counterexample, as it is not within the claimed concept definition. Rather, the object is seen as a monster which should not be allowed to disrupt a harmonious theorem. For instance, one of the students suggests that the hollow cube (a cube with a cube-shaped hole in it) is a counterexample to Euler's conjecture, since $V - E + F = 16 - 24 + 12 = 4$. Another student uses monster-barring to argue that the hollow cube does not threaten the conjecture as it is not in fact a polyhedron. The concept polyhedron then becomes the focus of the discussion, with the definition being formulated explicitly for the first time; as 'a solid whose surface consists of polygonal faces' (according to which, the hollow cube *is* a polyhedron), and 'a surface consisting of a system of polygons' (according to which, the hollow cube is *not* a polyhedron) [23, p. 14]. Using this method, the original conjecture is unchanged, but the meaning of the terms in it may change. Monster-adjusting is similar, in that one reinterprets an object in such a way that it is no longer a counterexample, although in this case the object is still seen as belonging to the domain of the conjecture. The example in [23] 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'.

Lemma incorporation works by distinguishing global and local counterexamples. The former is one which is a counterexample to the main conjecture, and the latter is a counterexample to one of the proof steps (or lemmas). A counterexample may be both global and local, or one and not the other. When faced with a counterexample, the first step is to determine which type it is. If it is both global and local, *i.e.*, there is a problem both with the argument and the conclusion, then one should modify the conjecture by incorporating the problematic proof step as a condition. If it is local but not global, *i.e.*, the conclusion may still be correct but the reasons for believing it are flawed, then one should modify the problematic proof step but leave the conjecture unchanged. If it is global but not local, *i.e.*, there is a problem with the conclusion but no obvious flaw in the reasoning which led to the conclusion, then one should look for a hidden assumption in the proof step, then modify the proof and the conjecture by making the assumption an explicit condition.

Proofs and refutations consists of using the proof steps to suggest counterexamples (by looking for objects which would violate them). For any counterexamples found, it is determined whether they are local or global counterexamples, and then lemma incorporation is performed.

6. Walton's argumentation schemes

Walton is an argumentation theorist and informal logician, with a particular interest in legal argumentation and AI. He and his colleagues have catalogued a hundred or so argumentation schemes: "forms of argument that model stereotypical patterns of reasoning" [56, 195]. This underlying idea has a lengthy history, dating back ultimately to Aristotle's *Topics*, through a 'topical tradition' including Cicero, Boethius, and many mediæval logicians, before eventually falling into disuse in the early modern era [54, p. 275 f.]. This tradition has been resuscitated by modern informal logic. Several subtly different characterizations of the argumentation scheme (or argument scheme) are now in use. We shall investigate what has become the most influential treatment, that of Walton and collaborators, which has been defended in several books and articles since the early 1990s.

Walton's argumentation schemes are presented as schematic arguments, typically accompanied by *critical questions*. The critical questions itemize known vulnerabilities in the argument, to which its proposer should be prepared to respond. In principle, they can always be incorporated into the schematic argument as additional premisses [54, p. 17]. This move has advantages in formal implementations of argumentation schemes, but risks obscuring the dialectical context in which the schemes are characteristically employed. Many of the schematic arguments are special cases of *modus ponens* in which the hypothetical premise lacks the force of a deductive implication. Hence most of Walton's argumentation schemes are presumptive or defeasible, although deductive inferences can also be understood as argumentation schemes.

Part II. Some salient connections

7. Abduction and other dangerous things

7.1. Abduction and mathematics. Peirce was a mathematician in his own right, with a whole volume of collected papers devoted to mathematics [39, vol. 4], developing notions of the infinite, continuity and real numbers independently of Dedekind and Cantor (compared in [12]). (His contributions to mathematics have been highlighted by Carolyn Eisele: see, in particular, her edition of [40].) It is somewhat surprising therefore, that there is no indication that Peirce saw abduction as a method to be employed in mathematics, reserving it for scientific discovery and explicitly saying that "Deduction must include every attempt at mathematical demonstration" [39, 5.590], and that "Deduction is the only necessary reasoning. It is the reasoning of mathematics" [39, 5.145]. In spite of this, we strongly believe that Peirce's abduction can be usefully applied to mathematical thinking.

There are several candidates for what would constitute an initial, surprising, fact in mathematics. The most obvious would be a conjecture, C, where the explanation for C would be a proof plan A. This raises the question of what comes first, a theorem (conjecture) or its proof (or proof idea). In some cases it seems that a proof plan came *before* a conjecture, such as Euler's proof of the Basel problem. This is the problem of finding the sum of the reciprocals of the squares of the natural numbers.[6] Euler drew an analogy between finite and infinite series, and applied a rule about finite series to infinite series to find his conjectured value of $\frac{\pi^2}{6}$, *i.e.*, he calculated the value (found the conjecture) *during* the proof attempt. However, there are many other cases in which a conjecture exists without proof or even a proof idea: this is shown by mathematicians having an intuition about (either belief in, or scepticism of) open conjectures. For example, the four colour theorem was widely believed to hold *before* it had been proved; likewise $P = NP$, although still open, is widely held to be false. Thus, a mathematician may well be in the position of trying to explain an initial 'fact'. Lakatos discusses the "heuristic precedence of the result over the argument" in his first footnote on [23, p. 9], citing both modern and ancient mathematicians who hold that "[...] it is [...] necessary to know beforehand what is sought" ([21, Bk I, p. 129], cited on [23, p. 9]; see also [49] for discussion on this issue). Mancosu [33] discusses ways in which mathematical activity mirrors the occurrence of explanatory hypotheses in science, where mathematics is seen as hypothetico-deductive in nature. Lakatos is one such writer who views mathematics in this way, thus the overlap with Peirce's development of abduction: other writers include Mill, Russell and Gödel. Leng [31, p. 105] also argues that some mathematical proofs are explanatory and discusses whether theories of scientific explanation, in particular inference to the best explanation, apply to mathematical explanation.

7.2. Surprise in mathematics. The role of surprise in mathematics mirrors that in science and other areas of discovery or invention. It might be thought to lead to deeper theorems or more important revisions if the counterexamples are surprising. Surprisingness in mathematics has been used as a way of evaluating concepts and conjectures in automated theory formation programs. Colton [10, p. 17] describes Epstein's Graffiti [15], which discards conjectures it deems unsurprising and his own HR system [9], which can be set to prioritise surprising conjectures, where surprisingness is based on semantic information about the concepts in-

[6] That is, the exact value of the infinite series $1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \frac{1}{36} + \ldots$. (see [50, p. 157–165] and [14, pp. 39–60] for commentary).

volved. He also describes Lenat's AM program [30], which measures the surprisingness of concepts (a concept is considered to be more interesting if it has a property which its parents do not have).

7.3. Abducting proofs and refutations. With our Lakatosian spectacles on, the obvious initial, surprising, fact would concern the existence of counterexamples.

> We observe a surprising fact, C, that there are counterexamples to our conjecture;
> But if we perform piecemeal exclusion/monster-barring/monster-adjusting/lemma-incorporation producing a modified conjecture, A, then C would be a matter of course;
> Hence, there is reason to suspect that A is true (in this context 'useful' might be a better word).

As an example, following Lakatos's case study, piecemeal exclusion can be presented as the following abductive argument.

> The surprising fact, C, that "*a hollow cube breaks the original conjecture that for all polyhedra* $V - E + F = 2$" is observed;
> But if $A =$ "*for all polyhedra except those with cavities* $V - E + F = 2$" were true, then C would be a matter of course;
> Hence, there is reason to suspect that A is true.

Thus, different methods present different 'hypotheses', to be 'set down upon our docket'. We can apply Peirce's criteria for evaluating hypotheses to the possible responses. Firstly, they must account for the counterexamples just found. Secondly, we must be able to test the resulting modified conjecture, proof or concept definition (here 'testable' relates to some agreed subset of vocabulary or observations). Thirdly, the modification should be worth investigating: we should be able to prove or disprove it relatively easily, it should be the simplest explanation we can find, there shouldn't be any immediately obvious counterexamples to the new conjecture, and it should relate to other conjectures, concepts or proofs in our theory. These criteria are implicit in the discussion in [23], for instance, when the students discuss the value of a new concept definition. Abduction consists of generating different conjectures which explain (*i.e.*, allow us to deduce) the surprising fact that particular counterexamples break a previous conjecture, and deciding which of the conjectures are worth further examination. For strategic withdrawal, Peirce's abduction can be used to find an initial or a modified conjecture where the 'surprising facts' are actually the supporting examples. For hidden lemma-incorporation, the initial surprising fact might be that there is a counterexample which appears to be global but not local. Peirce's abduction might be used to identify a suitable assumption which a mathematician is making in the proof, make it explicit and use it to repair the conjecture. Note that for the method of proofs and refutations, the discovery itself usually suggests the explanation.

Whether mathematics fits in with Peirce's three-stage model of enquiry is another question: a possible breakdown in the Lakatosian style of reasoning might be to firstly use induction to find a naïve conjecture (Polya describes this process in terms of induction in [45, pp. 35–41]). The existence of counterexamples is then discovered. Secondly, abduction is used to find a useful repair, such as a likely conjecture, which explains the fact of a given counterexample breaking the naïve conjecture (according to our argument above). Thirdly, deduction is used to draw out and prove consequences of the new conjecture.

7.4. Peirce and Toulmin. Peirce's abduction could be used to *find* and elegantly state the conditions: for instance, rather than state the isolated exception "Harry is not a British subject", the rebuttal generalises to "both his parents were aliens". (This is exactly the same as going from the 'surprising' fact that "the hollow cube does not have an Euler characteristic of 2", to finding, evaluating and stating the general rule that Euler's conjecture holds "except for polyhedra with cavities".)

7.5. Walton's abductive schemes The Argumentation Scheme for Argument from Verification below is more or less affirming the consequent, and thereby Peirce's abduction.

7.5.1. Argument from Verification.
Premise If A (a hypothesis) is true, then B (a proposition reporting an event) will be observed to be true.
Premise B has been observed to be true, in a given instance.
Conclusion Therefore, A is true.
Critical Questions:

1. Is it the case that if A is true then B is true?
2. Has B been observed to be true?

3. Could there be some reason why B is true, other than its being because of A being true? [54, pp. 331 f.]

Walton also offers something more specific:

7.5.2. Backward Abductive Argumentation Scheme.

Premise 1 D is a set of data or supposed facts in a case.

Premise 2 Each one of a set of accounts $A_1, A_2, \ldots, A_n$ is successful in explaining D.

Premise 3 A_i is the account that explains D most successfully.

Conclusion Therefore, A_i is the most plausible hypothesis in the case.

7.5.3. Forward Abductive Argumentation Scheme.

Premise 1 D is a set of data or supposed facts in a case.

Premise 2 There is a set of argument diagrams G_1, G_2, $\ldots$, G_n, and in each argument diagram D represents premisses of an argument that, supplemented by plausible conditionals and other statements that function as missing parts of enthymemes, leads to a respective conclusion C_1, C_2, $\ldots$, C_n.

Premise 3 The most plausible (strongest) argument is represented by G_i.

Conclusion Therefore, C_i is the most plausible conclusion in the case.

Critical Questions:

1. How satisfactory is A_i itself as an explanation of D, apart from the alternative explanations available so far in the dialogue?
2. How much better an explanation is A_i itself as of D than the alternative explanations so far in the dialogue?
3. How far has the dialogue progressed? If the dialogue is an inquiry, how thorough has the search been in the investigation of the case?
4. Would it be better to continue the dialogue further, instead of drawing a conclusion at this point? [54, pp. 329 f.]

The first two questions seem to apply only to the first scheme, but (it appears that) the last two are intended to apply to both.

8. Laying out the layout

Toulmin initially considered his model to describe non-mathematical argument. For instance, he argued 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. [52, p. 127]

However, he later represented Theaetetus's proof that there are exactly five platonic solids in his layout [53]. Toulmin's argumentation structure can represent more complex mathematical proofs. For example, one of us has shown how the proof that there are irrational numbers α and β such that α^β is rational [1], and the classical proof of the Intermediate Value Theorem [2] may be so represented. In addition, Alcolea [3] has used Toulmin's argumentation structure to represent meta-level mathematical argument, modelling Zermelo's argument for adopting the axiom of choice in set theory. Alcolea also presents a case study of Appel and Haken's computer assisted (object level) proof of the four colour theorem. (An alternative representation of this theorem, which also uses Toulmin's layout, is suggested in [1]). In [2], one of us describes ways of combining Toulmin's layout to represent yet more complex arguments, including the embedding of one layout within another, which is used to represent the proof that every natural number greater than one has a prime factorisation.

8.1. Toulmin and Peirce. At first reading, Toulmin's layout seems intuitive. However, even the process of representing deductive, inductive and abductive argument in Toulmin's layout requires some thought and led us to alternative representations. In Figure 2, we show two interpretations of deductive argument in Toulmin's layout: interpretation (a) follows the form of Toulmin's example arguments, for instance the Harry from Bermuda example; interpretation (b) is suggested by the way in which Toulmin describes the components of his layout, in particular the datum are the facts that we appeal to as a foundation for the claim and the warrant comprises rules, principles, inference-licences, propositions which show that the step from data to claim is legitimate [52, p. 98]. For inductive argument, shown in Figure 3 we show both the general and one specific case (interpretations (a) and (b) respectively). Toulmin describes Newton's ideas on "using our observations of regularities and correlations as the backing for a novel warrant" [52, p. 121], which suggests that interpretation (b) is closer to what Toulmin had in mind.

In Figure 4 we show two ways of using Toulmin's layout to represent abductive argument. Toulmin distinguishes between warrant-using and warrant-establishing arguments [52, p. 120], where the former is an argument in which the data is used to establish some conclusion by appealing to some accepted warrant, and the latter is an argument in which the acceptability of a new warrant is established by employing it in a number of cases in which both data and conclusion have been independently verified. In the warrant-establishing argument then, it is the warrant, rather than the conclusion, which is novel and "on trial". This latter type of argument sounds very much like abduction, and corresponds to interpretation (a) in Figure 4, where the backing might comprise the other arguments in which this warrant has been employed.

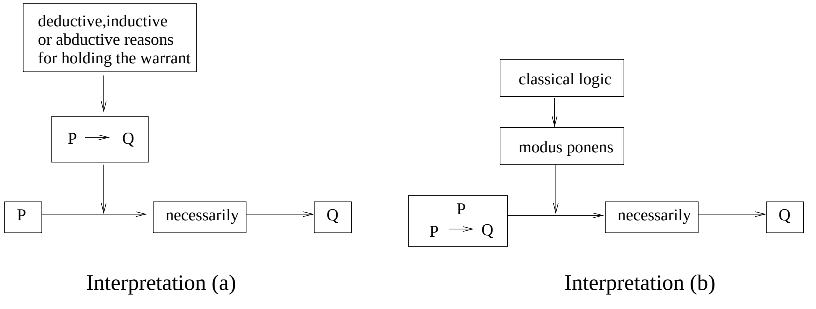

Figure 2. Two interpretations of deductive argument in Toulmin's layout

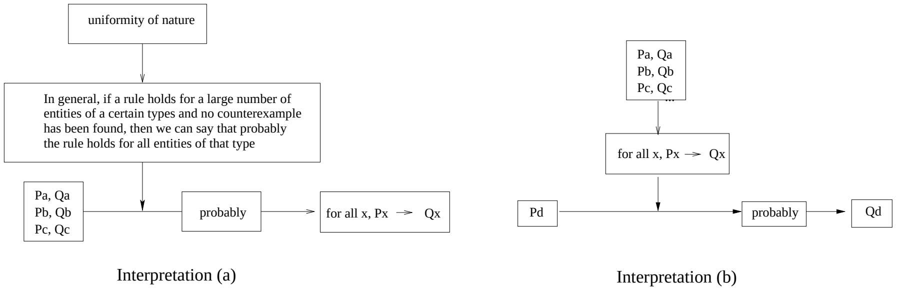

Figure 3. Using Toulmin's layout to represent: (a) general and (b) specific inductive argument

8.2. Toulmin and Lakatos. In figures 5 and 6 we show examples of how Toulmin's layout can be used to represent different levels of Cauchy's proof of Euler's conjecture in Lakatos's [23]. Note that one attractive as-

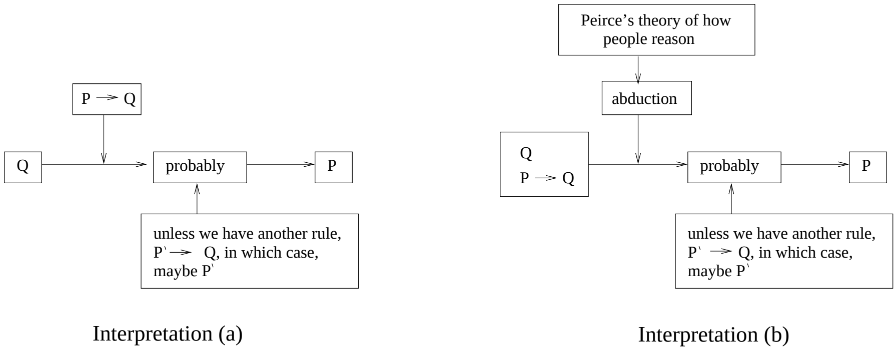

Figure 4. Two interpretations of abductive argument in Toulmin's layout

pect of Toulmin's layout is that it can be recursive: the backing, warrant, data and rebuttal can all themselves be the claim in another argument (part of the warrant in Figure 6 is the claim in Figure 5).

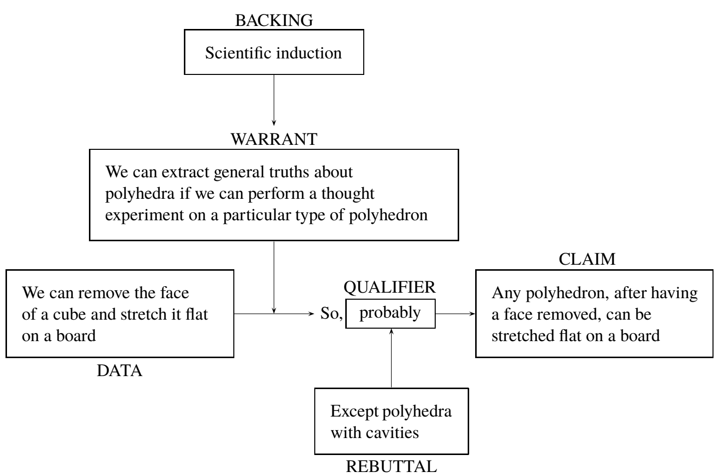

Figure 5. Example of how the first step of Cauchy's proof of Euler's conjecture can be represented in Toulmin's layout.

However, the roles that each aspect of the layout play can be ambiguous, and it is possible to express the same argument in multiple ways. All of Toulmin's examples in [52, Ch. 3] concern claims about specific "facts", supported by specific facts (data) and general principles (the warrant) which may themselves be supported (the backing)

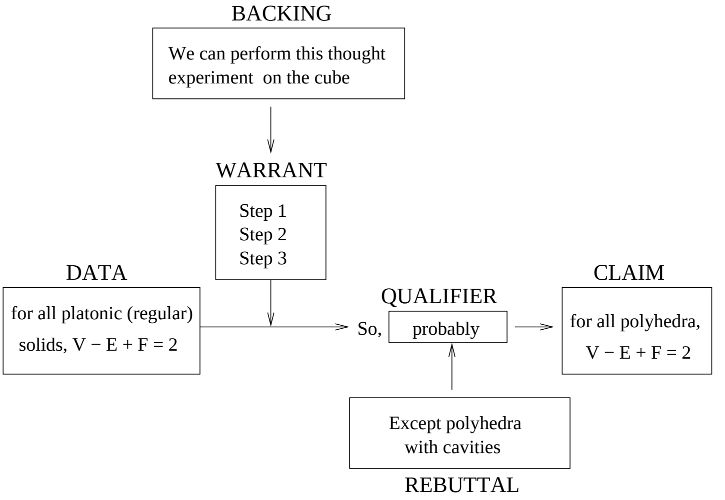

Figure 6. A representation of Cauchy's proof of Euler's conjecture, using Toulmin's layout, where steps 1–3 are as described in [23, p. 7], taking into consideration the first counterexample. The data are the facts which initially inspire the conjecture in Lakatos's 1976.

or may have exceptional conditions (the rebuttal). Toulmin gives the following example claims: Harry's hair is not black (p. 97 *ibid.*); Wilkins has committed an offence against the Road Traffic Acts (p. 97 *ibid.*); Peterson is not a Roman catholic (p. 111 *ibid.*); Jack has difficulty in walking (p. 115 *ibid.*); Anne has red hair (p. 124 *ibid.*); and, in applied mathematics, "on level ground this wall will cast a shadow of depth $10\frac{1}{2}$ ft.", based on *data* the height of the wall is 6 ft. and the sun is at an angle of elevation of 30°, the *warrants* are methods of geometrical optics (and the *backing* presumably is proof of these methods) (p. 137 *ibid.*). These examples suggest that Toulmin intended the input/output (data/claim) relation to be specific facts and the warrant the bridge between the two. In this case, the claim in Figure 6 becomes the warrant. In Figure 7 we represent the arguments surrounding Cauchy's conjecture in a way more faithful to Toulmin's own examples.

8.3. Toulmin and Pollock. In Figure 8 we show how Pollock's undercutting example might be represented in terms of Toulmin's layout. Given the datum that an object x looks red to me, I use the warranted rule that if x looks red to me then I have reason to believe that it is red (backed up by the rule that perception sensors are normally reliable)

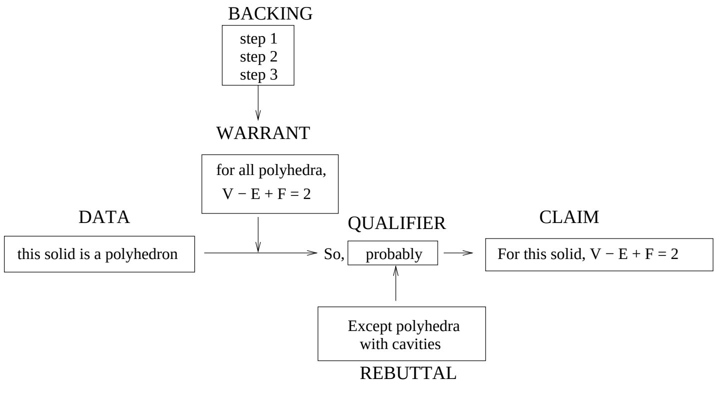

Figure 7. A different representation of the work discussed in [23] in Toulminian terms.

to conclude that probably, unless x is illuminated by red lights, x is red. We further discuss the relationship between Toulmin's rebuttals and Pollock's rebutters and undercutters in §12.3.

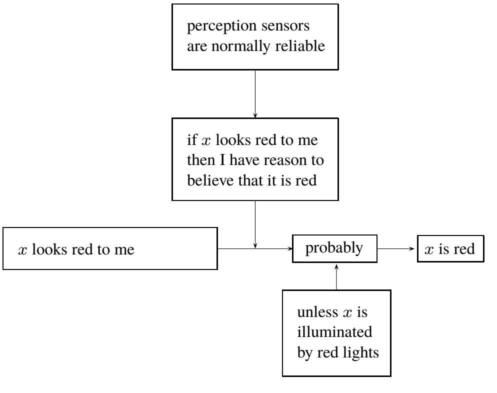

Figure 8. A representation of Pollock's undercutting example in terms of Toulmin's layout.

8.4. Toulmin and Walton. There is at least a family resemblance between argumentation schemes and Toulmin layouts. Most schemes could be expressed as layouts: the data and warrant can be seen as corresponding to premisses,[7] the claim (suitably modified by the qualifier) to the con-

[7] Some argumentation theorists, however, such as Hitchcock [22], have argued that a warrant is *not* a premise, and that Toulmin's introduction of the new termi-

clusion, and the backing and rebuttal both comprise possible answers to critical questions. However, since schemes are generic and layouts are (typically) specific (discussed in §8.2), it may be more productive to see layouts as corresponding to instantiations of schemes. The translation seems straightforward, at least when going from a layout to a scheme. Going the other way clearly poses some questions: Which premisses are data and which warrant? Which is claim and which is qualifier? Which critical questions are backing and which rebuttal? All of these arise in any application of layouts, not just to instantiations of schemes. The first question is a known problem for the application of Toulmin; we also address this in our figures 2, 3 and 4. Rather facetiously, we may answer the second question by saying that the qualifier is usually an adverb! However, there is a genuine problem of what to do when qualifiers are absent. As Matthew Inglis points out [personal communication], this is one of the weaknesses of the scheme approach, but could presumably be remedied by insisting on an explicit qualifier in every scheme. The third question is tougher. The difference between backing and rebuttal may track that between critical questions where the burden of proof is on the proponent or respondent, respectively (a distinction explored in [56, p. 209]). So the backing would comprise satisfactory answers to critical questions which the proponent ought to have worked out in advance, and the rebuttal defeating answers to critical questions the proponent need not have prepared against. Consider the following passage:

> The trustworthiness and consistency critical questions seem to have a positive burden of proof attached to the side of the questioner. The other [expertise, field, opinion, backup evidence] critical questions can just be asked out of the blue, so to speak. Once asked, this type of critical question must be given an appropriate answer or the original argument falls down. With these critical questions, the burden of proof remains on the side of the proponent of the appeal to expert opinion. Merely asking the question makes the original argument default. Asking the trustworthiness or consistency critical questions is a harder task.
> [56, p. 209]

---

nology improved on the traditional major premise-minor premise framework in terms of support for conclusions. Thus, Hitchcock argues, blurring the distinction between warrant – something in accordance with which the conclusion follows – and premise – something from which the conclusion follows – would be a step backward: “A warrant is an inference-licensing rule, not a premise” [22, p. 71].

Walton and Reed go on to say:

> According to this solution, [. . .] [expertise, field, opinion, and backup evidence, where the burden is on the proponent] for the appeal to expert opinion attack some aspect of one of the premisses in a way that *undercuts* the inference structure on which the argument is based. The remaining two critical questions bring in additional assumptions on which the argument is based, but at a deeper level. It is more of a *background assumption* that the expert is trustworthy, and is not biased, at least too heavily. And it is also more of a background assumption that what the expert says is consistent with what the other experts in the field say. Both assumptions can be violated and the argument may not be too badly off. But if either can be backed up strongly enough, it can certainly attack and destroy the original argument.
>
> [56, p. 210, our emphasis]

This is terminologically unhelpful, since Walton and Reed talk of undercutting not by a rebuttal but by the absence of backing, and of background assumptions not in the backing but as contradicting the rebuttal. But we think it makes sense. The absence of backing undercuts one of the premisses (the warrant, specifically); a successful rebuttal contradicts an unstated background assumption.

One solution may be to argue that backing and rebuttal are interchangeable, or rather complementary: the backing asserting that the rebuttal does not apply. However, this is problematic. For instance, in the case of Harry and Bermuda, as the statutes have changed since 1958, that may undercut the claim, but 'unless the statutes have changed' does not seem like a rebuttal.[8] Conversely, 'Harry's parents were not aliens' has no place in the backing, as it does not support the warrant.

9. Lakatos and his refutations

9.1. Lakatos and Peirce. In terms of background, Lakatos is closest to Peirce: both were describing techniques for discovery, both emphasised the importance of guessing, both ascribed importance to the role of surprise in discovery and, while Peirce was describing heuristics that people use for scientific discovery and Lakatos the mathematical case,

[8] The statutes have certainly changed. By virtue of the British Nationality Act 1981 and the British Overseas Territories Act 2002, Harry would now be a British Overseas Territories citizen, not a British subject.

Lakatos (better known for his philosophy of science than his mathematics) argued that "Mathematical heuristic is very like scientific heuristic — not because both are inductive, but because both are characterised by conjectures, proofs and refutations" [23, p. 74]. In a footnote Lakatos praises Polya for his "correct vision of deep analogy between scientific and mathematical heuristic" [*ibid.*, p. 74], criticising him only for his view that science, and therefore mathematics, is inductive: "... being indoctrinated that the path of discovery is from facts to conjecture, and from conjecture to proof (the myth of induction), you may completely forget about the alternative: deductive guessing." [*ibid.*, p. 73]. This sounds very much like Peirce.

9.2. Lakatos and Walton We see many of the undercurrents of Walton's schemes flowing through the discussion in [23], such as his *Pop Scheme* [54, p. 311], in which pressure from a particular group to accept a conclusion leads to its acceptance: this mirrors the atmosphere in Lakatos's discussion, in which members of the group all accept a specific definition, for the sake of the argument. (Walton doesn't specify that the accepter be a member of the group, although this is the case in [23]; another difference is that Walton is describing acceptance of conclusions, Lakatos acceptance of definitions.) Walton's *Ethotic Argument* [*ibid.*, 41, pp. 336 f.] could be seen reflected in the authority which Lakatos grants the teacher in his discussion, whose extra prestige as compared to that of the students leads his/her words to carry more weight (this varies from Walton's scenario in that the extra weight is granted by authority as opposed to moral character, but is the same principle). Walton's *Argument from Analogy* [*ibid.*, 7, pp. 315 f.] could be seen as describing the initial formulation of Euler's conjecture, in which a relationship between edges and vertices is noticed in two-dimensional shapes, and thus it is conjectured that an analogous relationship may exist for three-dimensional shapes ([23, p. 6], see also [37]). Indeed, one could re-write much of the discussion in [23] in terms of Walton's schemes.

However, the schemes in Walton's catalogue do not describe Lakatos's explicit and most important methods. The main difference is that the conclusions of almost all of Walton's schemes (excepting only *Argument from Precedent* [*ibid.*, 57, pp. 344 f.], described in Section 12.4.2 below) result in a recommendation for acceptance or rejection of a conclusion (some with qualification, for instance, "there is reason in favour of A" [*ibid.*, 4, pp. 331 f.]). No suggestions are made for what to do next;

in particular, how to modify an argument that one has rejected. Conversely, Lakatos specialises in heuristics which describe ways in which flawed arguments may be repaired.

In Sections 10 to 13 we consider both inter- and intra-theory connections between Lakatos's theory and other theories of argumentation.

10. Lakatos's surrender method

While falsification of hypotheses might be thought more characteristic of empirical science, Lakatos's account of the development of the Euler Conjecture describes exactly this type of reasoning, leading to his view that some aspects of mathematics proceed in a quasi-empirical fashion [23, p. 5]. He considers his method of surrender, to abandon a conjecture when faced with a counterexample ("Scrap the false conjecture, forget about it, and try a radically new approach" [*ibid*, p. 13]) to be the least productive of those he identifies. This method is Popper's naïve falsificationism applied to mathematics, and corresponds to Pollock's rebutting defeater, and to Walton's scheme for Argument from Falsification. (Together with the *Argument from Verification* shown above (§7.5.1) these schemes comprise *Argument from Evidence to a Hypothesis.*)

Argument from Falsification.
Premise If A (a hypothesis) is true, then B (a proposition reporting an event) will be observed to be true.
Premise B has been observed to be false, in a given instance.
Conclusion Therefore, A is false.

Critical Questions:

1. Is it the case that if A is true then B is true?
2. Has B been observed to be false?
3. Could there be some reason why B is true, other than its being because of A being true? [54, pp. 331 f.]

Let A be Euler's conjecture and B the proposition "the hollow cube is Eulerian", then the scheme is simple *modus tollens*. Argument from Verification is the chain of reasoning used for supporting examples. Of the critical questions for Argument from Falsification, negative responses to CQ1 correspond to monster-barring and to CQ2 to monster-adjusting (we describe both in §11). A negative response to CQ3 does not seem to have an analogue in [23].

11. Lakatos's definition-changing methods

Definitional change can occur within an argument, for instance, an argument in which a concept is defined differently in the premisses than in the conclusion, or the extension of a concept is unclear. Analysis of such changes dates back at least to Socrates and Plato's development of the concept of justice in *The Republic*.[9] This is one of the cornerstones of Lakatos's work, reflected in his methods of monster-barring and monster-accepting (subsequently developed into monster-*embracing* or *accepting*) in which borderline cases can be used to further understanding of a concept. It is perhaps surprising that neither Peirce, Pollock nor Toulmin consider ambiguity of terms, and that only two of Walton's schemes address anything approaching it: *Argument from Vagueness of a Verbal Classification* and *Argument from Arbitrariness of a Verbal Classification* (sections 11.2.1 and 11.3.1 below).

11.1. The monster-barring method. Lakatos's monster-barring method exploits any ambiguity in concepts, in order to defend a conjecture. In the first statement of Euler's conjecture, that for all polyhedra, $V - E + F = 2$, it is assumed that the extension of polyhedron is known, *i.e.*, we can distinguish between objects which are and are not polyhedra, even if the definition is not explicitly agreed. Once an object of ambiguous status arises, students do explicitly define polyhedra, *i.e.*, they start with a vague definition and make it more specific; although the new definition may include ambiguous sub-concepts such as *polygon*, *area*, and *edge*, whose definitions are also open to debate. The only criterion for a candidate definition is that it distinguish the agreed polyhedra from agreed non-polyhedra. The method of monster-adjusting also exploits

[9] Simonides proposes that "it is right to give back what is owed" [41, pp. 8–9]. This initial statement is questioned by Socrates with the counterexample of someone borrowing weapons from a friend who subsequently goes insane, in which case it would not be right to return the weapons. The discussion in [41] then turns to what it means to give back what is owed, with Polemarchus suggesting that people owe their friends good deeds, and their enemies bad ones. The dialogue later turns to what the concept of *doing right* means, and leads into Plato's treatment of justice. Larvor [29] argues that this discussion is an example of Lakatos's monster-barring.

ambiguity in concepts, but reinterprets an object in such a way that it is no longer a counterexample. The example in [23] concerns the star polyhedron, which 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 an Euler characteristic of -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 an Euler characteristic of 2. The argument then turns to the definition of 'face'. Monster-adjusting can be seen as a type of monster-barring, where the concept in question may be a concept in the consequent of an implication or equivalence conjecture, rather than the domain. If we formalise monster-barring as follows: from conjecture $\forall x(Px \rightarrow Qx)$, and (known) counterexample m such that Pm and $\neg Qm$, (re)define P as P' so that for the m in question, $\neg P'm$. Monster-adjusting can be seen in terms of this formalisation, where Q is (re)defined as Q' in such a way that $Q'm$.

11.2. Ambiguity in mathematics. The existence of ambiguity can be a legitimate critique of mathematical argument. Lakatos's work has demonstrated how much mathematics can be understood as progressively sharpening the initial vagueness of terms in a mathematical problem or hypothesis. Increases in rigour might often be understood in such terms.[10] Conversely, vagueness can vitiate mathematical discourse irreparably: consider Hilbert's Fourth Problem, read at the International Congress of Mathematicians in 1900, to *construct and study the geometries in which the straight line segment is the shortest connection between two points*, in which notions of distance, or metrics (and thus the meaning of *shortest*) are so vague that the problem is considered insolvable in its given form [17, p. 287]. (Nonetheless, much interesting mathematics has been developed from this problem, which was at least in part Hilbert's aim.) Another example can be found in Hilbert's sixth problem: *to treat in the same manner, by means of axioms, those physical sciences in which mathematics plays an important part; in the first rank are the theory of probabilities and mechanics.* The scheme for Argument from Vagueness of a Verbal Classification is a step in the right direction, but

[10] Grosholz [18] discusses ambiguity in mathematics, using historical case studies of language and notation in areas of mathematics including group theory, geometry, mechanics, topology and set theory, to show that representation and ambiguity play key roles in mathematical discovery, and arguing that various advances have been made *because* of the irreducible ambiguity in notations and diagrams.

is still not sufficient, not least because the only response to vagueness suggested is rejection of the argument. Presumably the proponent then advances a less vaguely expressed argument, but Walton et *al.* make no suggestion as to how, as opposed to Lakatos's suggested heuristics.

11.2.1. Argument from Vagueness of a Verbal Classification.

Premise If an argument, Arg occurs in a context of dialogue that requires a certain level of precision, but some property F that occurs in Arg is defined in a way that is too vague to meet the requirements of that level of precision, then Arg ought to be rejected as deficient.

Premise Arg occurs in a context of dialogue that requires a certain level of precision that is appropriate for that context.

Premise Some property F that occurs in Arg is defined in a way that is too vague to meet the requirements of the level of precision appropriate for that context.

Conclusion Therefore, Arg ought to be rejected as deficient.

Critical Questions:

1. Does the context of dialogue in which Arg occurs demand some particular level of precision in the key terms used?
2. Is some property F that occurs in Arg too vague to meet the proper level or standard of precision?
3. Why is this degree of vagueness a problem in relation to the dialogue in which Arg was advanced? (cf. [54, pp. 319 f.])

11.3. Arbitrariness in mathematics. It might seem to be a category mistake to criticise mathematics for its arbitrariness: at least on one understanding, *all* of mathematics is arbitrary. Arbitrariness is clearly a telling criticism of the *application* of mathematics. Consider the familiar case of a carefully worked out model of dubious fit to the phenomenon under investigation. Here the arbitrariness is in the assumptions that were made before the mathematics was begun. But Lakatos's complaints of *ad hoc*-ness, in monster barring for example, might also be understood as exemplifying this scheme. This criticism is reflected in the bitter complaining about definitions, for instance, "I admire your perverted ingenuity in inventing one definition after another as barricades against the falsification of your pet ideas." [p. 16], and definitional change called "linguistic tricks" [pp. 19, 20], "rescue-definitions" [p. 20], "weird shifts in meaning" [p. 21], "mutilation of concepts", [p. 22], and an explicit reference to "arbitrary definitions" [p. 50] (clearly this definitional change is not arbitrary with respect to the conjecture under discussion, rather the criticism is that the proposed new meanings are arbitrary in general). Certainly avoidance of arbitrariness in the latter sense is desirable, but again Walton offers no response other than rejection.

11.3.1. Argument from Arbitrariness of a Verbal Classification.

Premise If an argument, Arg occurs in a context of dialogue that requires a non-arbitrary definition for a key property F that occurs in Arg, and F is defined in an arbitrary way in Arg, then Arg ought to be rejected as deficient.

Premise Arg occurs in a context of dialogue that requires a non-arbitrary definition for a key property F that occurs in Arg.

Premise Some property F that occurs in Arg is defined in a way that is arbitrary.

Conclusion Therefore, Arg ought to be rejected as deficient.

Critical Questions:

1. Does the context of dialogue in which Arg occurs require a non-arbitrary definition of F?
2. Is some property F that occurs in Arg defined in an arbitrary way?
3. Why is arbitrariness of definition a problem in the context of dialogue in which Arg was advanced? (cf. [54, p. 320])

11.4. Methodology in mathematics. The section on monster-barring in [23] is interspersed with heated debate on methodology: whether mathematicians should study typical, ordinary examples and generate interesting and useful theorems about these, or focus on boundary cases, studying mathematics in its "critical state, in fever, in passion" [23, p. 23]. The teacher concludes that monster-barring is *not* a valid method; indeed, it is presented as the least sophisticated method after the method of surrender. The main criticisms are that monster-barrers are anti-falsificationists who defend a conjecture at any cost, which makes the conjecture deteriorate into meaningless dogma, and that the method is *ad hoc*, since the border between monsters and counterexamples is done in fits and starts. "Using this method one can eliminate any counterexample to the original conjecture by a sometimes deft but always *ad hoc* redefinition of the polyhedron, of its defining terms, or the defining terms of its defining terms. We should somehow treat counterexamples with more respect, and not stubbornly exorcise them by dubbing them monsters" (*Teacher*, in [23, p. 23]). The Duhem-Quine thesis [46], that a scientific theory cannot be tested in isolation since a test of one theory always depends on other assumptions or hypotheses, is also relevant to this discussion. We cannot falsify a conjecture, rather we can show that a collection of assumptions, concepts, counterexample and conjecture is internally inconsistent. The choice then arises as to which of the collection we reject. Monster-barrers would argue that we should reject the counterexample and certain concept definitions, and retain the conjecture, whereas critics of monster-barring might argue that we should reject the conjecture under discussion.

12. Lakatos's conjecture-changing methods

12.1. Lakatos's exception-barring methods. Lakatos's exception-barring methods, piecemeal exclusion and strategic withdrawal, are presented early on in [23, pp. 24–30]; a sign that they are considered unsophisticated and inferior to later methods. However, they are worth considering in some detail, because their legacy can be seen throughout the book: his later methods of lemma incorporation are essentially proof-oriented versions of exception-barring. This is made explicit by Lakatos in his distinction between "primitive exception-barring", which makes no use of a proof, and lemma-incorporation:

> The best exception-barrers do a careful analysis of the prohibited area... in fact your method [the method of lemma-incorporation] is, in this respect, a limiting case of the exception-barring method [. . .].
>
> [23, p. 37]

The exception barring methods target propositions which are "hopefully false" (as opposed to ones which are true — which should be accepted, or "hopelessly false" — which should be rejected — p. 26, *ibid.*), *i.e.*, propositions which hold for most but not all examples considered. They consist in determining the domain of validity for a claim, and result in a modified version of a conjecture (or claim). This may be done by identifying specific counterexamples, generalising from these to form a class of exceptions and excluding this class in a Toulminian fashion (Toulmin's rebuttal seems to be written with exactly this situation in mind; stating the conditions under which a claim does not hold). A second approach is to "withdraw to safety", by generalising from some specific supporting examples and limiting the claim to only this class.

Both methods are criticised as resulting in a "chaotic position" (p. 25, *ibid.*), since we can neither know when all exceptions have been identified and safely excluded, nor even whether some may lie within our supposedly safe stronghold. The latter method is further criticised as even if the boundaries are drawn so narrowly as to be certain of the truth of our proposition, if there are further supporting examples which lie outwith the boundaries then we have not succeeded in our task: the object of the exercise is to draw the *right* boundaries — as wide as possible and no wider — rather the *safest* boundaries, which result in a dull conservatism and a failure to illuminate.

12.2. A logical representation of exception-barring. Logically, these methods can be expressed as follows: given conjecture $\forall x(Px \rightarrow Qx)$, a set of counter (or negative) examples Neg such that $\forall n \in Neg(Pn \wedge \neg Qn)$, and a set of positive examples Pos such that $\forall p \in Pos(Pp \wedge Qp)$;

(i) find a concept C_1 such that for all n, $C_1 n$, and for all p, $\neg C_1 p$, and modify the conjecture to $\forall x((Px \wedge \neg C_1 x) \rightarrow Qx)$ (piecemeal exclusion), and

(ii) find a concept C_2 such that for all p, $C_2 p$, and for all n, $\neg C_2 n$, and modify the conjecture to $\forall x((Px \wedge C_2 x) \rightarrow Qx)$ (strategic withdrawal).

These two methods are logically equivalent if $\forall x(Px \rightarrow (C_1 x \vee C_2 x))$. For instance, given the property of being an integer (P), the property of being a number with an even number of divisors (Q), and an initial conjecture: $\forall x(Px \rightarrow Qx)$, this could be modified by examining counterexamples (1,4,9,16, ...), generalising from them to the property of being a square number (C_1), and excluding this class from the domain to get $\forall x((Px \wedge \neg C_1 x) \rightarrow Qx)$ (all integers except squares have an even number of divisors). Alternatively, the same conjecture could be formed by using strategic withdrawal to generalise from supporting examples (2,3,5,6,...) to the property of being a non-square number (C_2), resulting in the modified conjecture $\forall x((Px \wedge C_2 x) \rightarrow Qx)$ (all non-square integers have an even number of divisors). This possibility for logical equivalence is not noted by Lakatos, maybe because none of his examples of exception-barring result in two concepts where one is the complement of the other. The fact that this is an equivalence theorem in mathematics, *i.e.*, $\forall x((Px \wedge C_2 x) \leftrightarrow Qx)$ (x is a non-square integer if and only if it has an even number of divisors) also raises the question of how these methods apply to types of conjecture other than implications.

Again, this is not considered by Lakatos, and is also often overlooked in argumentation literature: all examples of Toulmin's claims are of the type Pa, and his warrants are generally implications $\forall x(Px \to Qx)$.[11]

12.3. A mediæval treatment of defeasibility. Syncategorematic terms were a preoccupation of many scholastic logicians from the twelfth century until the eclipse of traditional logic in the early modern era. These are terms which have no non-logical sense but which affect the logical form of a proposition. Reflection on these terms led to a doctrine of *exponible* propositions: propositions which could be analysed as conjunctions or disjunctions of simpler propositions. The scholastic analysis of *exceptive* syncategorematic terms, such as "besides", foreshadows the twentieth century debate over defeasibility which we have been discussing. Exceptive propositions were analysed as possessing at least two parts. For example, Walter Burley, writing in the early fourteenth century, states that:

> [E]ach exceptive has two exponents, an affirmative one and a negative one. For example, 'Every man besides Socrates runs' is expounded like this: 'Every man other than Socrates runs and Socrates does not run'. And 'No man besides Socrates runs' is expounded like this: 'No man other than Socrates runs and Socrates runs'. [7, p. 256]

(Some authors would add a third exponent 'Socrates is a man' to each pair [5, p. 235].) However, on most modern analyses, in each case only the first exponent is a logical consequence of the original statement. The second exponent would be seen as an implicature at best.

This historical detour may clarify the relationship between Toulmin's rebuttals and Pollock's rebutters and undercutters. Toulmin's rebuttals are introduced with an exceptive term, typically "unless". If this is understood as modifying the warrant in his layout (something on which there is less than complete consensus — see §12.5), then the warrant would be an exceptive proposition. For instance, the warrant in the Harry from Bermuda example would read 'anyone born in Bermuda will generally be British, unless his parents were aliens'. This would characteristically translate into first order logic as $\forall x((Bx \wedge \neg Ax) \to Sx)$, where $Bx =$ 'x was born in Bermuda', $Sx =$ 'x is a British subject', and $Ax =$ 'x's parents were aliens'. However, on Burley's reading, the warrant would

[11] Gasteren [16] addresses this imbalance, focussing on equivalence conjectures, and showing how analysis of the form of a conjecture can guide the design of its proof.

translate as $\forall x((Bx \wedge \neg Ax) \rightarrow Sx) \wedge \forall x(Ax \rightarrow \neg Sx)$, which entails $\forall x(Bx \rightarrow (\neg Ax \leftrightarrow Sx))$. Hence, on the narrower, modern reading of exceptive propositions, Toulmin's rebuttals are undercutters, but on the looser, scholastic reading they conjoin an undercutter and a rebutter. Toulmin preserves this ambiguity by eschewing such explicit formalisation, but most of the examples of rebuttals in [52] and [53] provide reasons to reject the corresponding claim, and would thus support the scholastic reading. Nonetheless, Toulmin does introduce some examples of rebuttals most plausibly understood on the modern reading, perhaps including Harry's parents being aliens, since that need not have prevented him from acquiring British nationality by naturalisation.

We can also fruitfully consider Lakatos's exception-barring methods in terms of exceptives, since the proposition resulting from piecemeal exclusion will be exceptive, whether affirmative, 'all polyhedra except polyhedra with cavities are Eulerian', or equivalently, negative 'no polyhedra except those with cavities are non-Eulerian'. In the discussion in [23] Lakatos implies that strategic withdrawal results in a negative exceptive 'no polyhedra but convex polyhedra are Eulerian'. (Again, this is equivalent to an affirmative exceptive, 'all polyhedra except convex polyhedra are non-Eulerian'). The discussion on boundaries suggests that despite the representation of the conjecture as an implication, the goal is to find an equivalence conjecture, *i.e.*, 'all polyhedra are Eulerian if and only if convex', which suggests that Lakatos was tacitly following the scholastic reading of his exceptive proposition. This is only explicitly discussed in the context of strategic withdrawal: "Could you have withdrawn too radically, leaving lots of Eulerian polyhedra outside the walls?" (p. 28, *ibid.*). Indeed, when the piecemeal-excluded conjecture is modified to *for all polyhedra that have no cavities (like the pair of nested cubes) and tunnels (like the picture-frame)* $V - E + F = 2$ (p. 26 *ibid.*), there is an obvious "counterexample" of a pyramid with a tunnel in it, for which $V - E + F$ *is* equal to 2 (shown in figure 9), of which no mention is made in [23]. Thus, in contrast with strategic withdrawal, Lakatos's description of piecemeal exclusion seems to focus on the logical implication, and thereby the modern reading of the exceptive.

12.4. Lakatos and Walton. In terms of Walton's argumentation schemes, his *Argumentation Scheme for Argument from an Exceptional Case* and *Argumentation Scheme for Argument from Precedent*, shown below, are both pertinent to Lakatos's piecemeal exclusion.

Figure 9. A pyramid with a tunnel in it. $V - E + F = 13 - 20 + 9 = 2$, and thus it is a supporting example for Euler's conjecture, however it is barred by the piecemeal-exclusion move to exclude all polyhedra with tunnels.

12.4.1. Argument from an Exceptional Case.

Premise For all x, if the case of x is an exception, then the established rule does not apply to the case of x.

Premise The case of a is an exception.

Conclusion Therefore, a need not do A.

Critical Questions:

1. Is the case of a a recognized type of exception?
2. If it is not a recognized case, can evidence why the established rule does not apply to it be given?
3. If it is a borderline case, can comparable cases be cited?

(cf. [54, p. 344])

Here we have an exception utilising step.

12.4.2. Argument from Precedent.

Premise The existing rule says that for all x, if x has property F then x has property G.

Premise But in this case C, a has property F, but does not have property G.

Conclusion Therefore, the existing rule must be changed, qualified, or given up, or a new rule must be introduced to cover case C.

Critical Questions:

1. Does the existing rule really say that for all x, if x has property F then x has property G?
2. Is case C legitimate, or can it be explained away as not really in violation of the existing rule?

3. Is case C an already recognized type of exception that does not require any change in the existing rule? (cf. [54, p. 344])

And here an exception establishing step. One can perhaps see how (an adaptation of) these schemes could be deployed to set up a system of defeasible rules, procedures or definitions. But we are not yet that close to Lakatos.

12.5. Revisiting Toulmin's rebuttal. Although perhaps seemingly obvious, there is room for ambiguity regarding whether Toulmin's rebuttal is intended to be general or specific. As discussed in §8.2, all of Toulmin's examples in [52, Ch. 3] concern claims about specific facts. This suggests to us that Toulmin intended the data and the claim to be specific facts and the warrant the (general) bridge between the two. In subsequent work however, such as [1, 2, 3] (discussed in §8), other authors have shown that Toulmin's layout can also be used for general claims. Toulmin's focus on the specific raises the question of whether he intended the rebuttal to be specific or general (this is equivalent to asking whether the rebuttal rebuts the claim or the warrant): in his examples it is specific. Consider his discussion of Anne's hair colour [52, p. 117]: based on the datum that *Anne is one of Jack's sisters*, and warrant that *any sister of Jack's may be taken to have red hair* (which itself has the backing that *all his sisters have previously been observed to have red hair*) we may conclude that, unless the rebuttal that *Anne has dyed her hair/gone white/lost her hair* holds, subject to the qualifier *presumably*, *Anne now has red hair*. Here, Anne *is named* in the rebuttal: this is one specific counterexample. We may have expected a general rebuttal, such as *any sister of Jack's, who has dyed her hair/gone white/lost her hair*; thus repairing the general warrant to: *any sister of Jack's who has not dyed her hair/gone white/lost her hair may be taken to have red hair*. The situation is particularly interesting in Toulmin's general cases, in which he uses a pronoun that might be reasonably taken to refer to the general *or* the specific case. For example, in "we can presumably claim that Harry is British, since anyone born in Bermuda will generally be British . . . , unless **his** parents were aliens", does "his" refer to Harry or to anyone born in Bermuda?

We see an analogue in the method of piecemeal exclusion, although not noted by Lakatos. In cases where there are few counterexamples (or only one), it may be preferable to exclude these by name, rather than generalising to a class and excluding that. For instance, in the conjecture *all primes are odd*, given the counterexample 2 we could generalise to the (singleton) class of smallest primes and modify our conjecture to *all primes except for the smallest one are odd*. However, possibly to avoid the extra inference step, this theorem is usually expressed as *all primes except 2 are odd*. Goldbach's conjecture, that *every even number except 2 can be expressed as the sum of two primes* provides a second example, in which the obvious classification "smallest even number" is passed over for the simpler "2". Indeed, examples of this type in mathematics usually, if not always, involve the smallest member of the domain of a claim, such as the trivial group, the empty set, the singleton graph, etc. A theorem which held for all primes except 287, or all groups except for the real numbers under addition, would be curious in the extreme and would certainly merit further investigation. In [36] one of us called this method "counterexample-barring".

13. Lakatos's proof-changing methods

13.1. Three types of lemma-incorporation. Lemma-incorporation is triggered by a counterexample, and comes in three flavours, depending on the type of counterexample. A *global counterexample* is a counterexample to the main conjecture, and a *local counterexample* is a counterexample to one of the proof steps (Lakatos calls these *lemmas*). In argumentation terminology (compare, §4), global counterexamples are rebutters for *any* argument that P, and local counterexamples are undercutters for (*some* step of) *some* argument that P. That is, a rebutter for an argument that P is (or implies the existence of) a derivation of $\neg P$; an undercutter for an argument that P is (or implies the existence of) a defeating answer for one of the critical questions in the derivation of P. Lakatos considers counterexamples which are both global and local, or one and not the other. He suggests that the first step, when faced with a counterexample, is to determine which type it is. If it is both global and local, *i.e.*, there is a problem with both the argument and the conclusion, then one should use strategic withdrawal to modify the conjecture, but — crucially — the domain to which we withdraw must be generated by the problematic proof step S_i. That is, we create a concept "all X which satisfy proof step S_i" and then limit the claim to this concept (in the Euler example it is "simple polyhedra", that is, polyhedra for which step 1 of Cauchy's proof can be performed [p. 34 *ibid.*]). This is a combination of both rebuttal and undercutting, where both of Pollock's defeaters come into play. If the counterexample is local but not global, *i.e.*, the conclusion may still be correct but the reasons for believing it are flawed, then standard piecemeal exclusion is used on the problematic proof step. That is, objects which support and refute the problematic proof step S_i are examined, a concept found which is true of all the supporting examples and false for all the counterexamples and the claim made in S_i then limited to all objects for which this concept holds (in the Euler example it is "boundary triangles", that is, step 3 of Cauchy's proof is modified to *removal of any* boundary *triangle preserves the Euler characteristic* [p. 11 *ibid.*]). The global conjecture is left unchanged. This is an example of just undercutting, which is captured by Pollock's undercutting defeater. If the counterexample is global but not local, *i.e.*, there is a problem with the conclusion but no obvious flaw in the reasoning which led to the conclusion, then Lakatos suggests searching for a hidden assumption in the proof, then modifying the culprit proof step and the global conjecture by making the assumption an explicit condition. This is a case of simple rebuttal, *i.e.* rebuttal without undercutting. Logically, we would have a global conjecture $\forall x(Px \to Qx)$, a set of counter (or negative) examples Neg such that $\forall n \in Neg\ (Pn \wedge \neg Qn)$, and a set of positive examples Pos such that $\forall p \in Pos(Pp \wedge Qp)$, and a set of conjectures which constitute proof steps for the global conjecture:

$$\begin{gathered}\forall x(P_1 x \to Q_1 x)\\ \forall x(P_2 x \to Q_2 x)\\ \vdots\\ \forall x(P_n x \to Q_n x)\end{gathered}$$

For one of the proof steps it may be possible to find a concept C_3 such that for all n, $C_3 n$, and for all p, $\neg C_3 p$. In this case the proof step should be modified to $\forall x(P_i x \to (Q_i x \wedge C_3 x))$ (a third form of exception-barring?): we would then have a global and local counterexample, which can be dealt with as discussed above.

13.2. A new, fourth type, of lemma-incorporation? While Lakatos does deal with the surprising case of a counterexample which appears to be global but not local, he does not consider the fourth scenario, in which we have an object which is neither a global nor local counterexample but still seems surprising and is not what was intended by the original claim. In this case we would expect to find a hidden assumption in the *global* conjecture. For instance, consider the beaker in Figure 10(a). This is Eulerian, since it has zero vertices, two edges and four faces, and, if we accept that the cylinder is not a counterexample to any of the proof steps then this is not either. However, it is (probably) not what was meant by the initial claim. Thus, while not a counterexample in the traditional sense, the beaker would constitute a fourth strange object for Lakatos (the same argument holds for the (topologically equivalent) bowler hat, in Figure 10(b).[12] The modification in this case could involve finding a hidden assumption in the global conjecture, making it explicit and then performing exception-barring (for instance, the criterion that the polyhedron must be a closed system). An alternative, since Lakatos would presumably prefer the method to involve the proof, would be to perform a version of global-only lemma-incorporation, in which an assumption hidden in the proof is identified, made explicit, and incorporated into the global conjecture (for instance, the criterion that it is possible to remove only one face at a time).

(a) The beaker (b) The bowler hat

Figure 10. Both the beaker and the bowler hat are Eulerian, since $V - E + F = 0 - 2 + 4 = 2$. However, they are (presumably) not the sort of objects intended to be covered by Euler's conjecture.

13.3. Rebutting without a known undercutter in mathematics. Lakatos's second case study follows the development of Cauchy's proof [8] of the conjecture that 'the limit of any convergent series of continuous functions is itself continuous' [23, App. 1]. He shows how hidden lemma-

[12] Figure 10(a) is cropped from a photo by AndreyTTL; Figure 10(b) is a photo by Fozrocket.

incorporation and the third kind of counterexample was used to repair the faulty conjecture and proof. The counterexample, found by Fourier is:

$$\cos x - \frac{1}{3}\cos 3x + \frac{1}{5}\cos 5x - \dots$$

which converges to the step function. Lakatos credits Seidel with inventing the method of proofs and refutations, arguing that he discovered it at the same time as he discovered the proof-generated concept of uniform convergence, and that he was aware of the importance of his method [23, p. 136]. He quotes Seidel:

> Starting from the certainty just achieved, that the theorem is not universally valid, and hence that its proof must rest on some extra hidden assumption, one then subjects the proof to a more detailed analysis. It is not very difficult to discover the hidden hypothesis.
>
> [51, p. 383], on [23, p. 136]

Given that in this example it took twenty six years after the proof was published (and thirty five years after Fourier's series became known) to identify the hidden assumption of uniform convergence in the proof, we may question how easy it is to discover a hidden hypothesis. Lakatos, however, suggested 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.[13] Lakatos's example of the cylinder in his main case study appears to have been concocted by himself, as a plausible

[13] This example is complicated: Cauchy's claim is generally regarded as obviously false, and the clarification of what was wrong is usually taken to be part of the more rigorous formalisation of the calculus developed by Weierstrass, involving the invention of the concept of *uniform convergence* (the historical route is sketched in [19, pp. 213–17]). The episode was treated by Lakatos in two different ways. Cauchy claimed that the function defined by pointwise limits of continuous functions must be continuous [8]. In fact, what we take to be counterexamples were already known when Cauchy made his claim, as Lakatos points out in his earlier analysis of the evolution of the ideas involved [23, App. 1]. After discussion with Abraham Robinson, Lakatos then saw that there was an alternative analysis. Robinson was the founder of non-standard analysis, which found a way to rehabilitate talk of infinitesimals, for example, positive numbers greater than zero, but less than any "standard" real number (see [48], first edition 1966). Lakatos's alternative reading, presented in [24], is that Cauchy's proof was correct, but that his notion of (real) number was different from that adopted by mainstream analysis to this day.

nineteenth century addition to the discussion, to show how an important method may have applied to his Eulerian example. (While [23] is a rational reconstruction, much of what the "students" discuss has some historical correlation, as evidenced by the numerous footnotes. The cylinder, first introduced by GAMMA [23, p. 22], has no associated historical footnote.[14]) The main discussion occurs on pp. 42–50, in which ALPHA puts forward various 'implicit assumptions' in Cauchy's proof, which the cylinder undercuts. Other students, especially GAMMA, argue that such assumptions have only just been invented, specifically for the purpose of being violated by the cylinder. In argumentation terms, ALPHA argues that any rebutter *must* also be an undercutter, although it may undercut a premise which is initially hidden or missing: proof analysis will make implicit undercutting explicit. GAMMA, on the other hand, argues that there *are* cases of mathematical counterexamples which are genuinely global but not local: that there can be rebutters without undercutters. The discussion culminates in the *Principle of Retransmission of Falsity* (p. 47); the criterion that a proof-analysis is valid and the corresponding mathematical theorem true if and only if there is no third type of counterexample, *i.e.*, if all global counterexamples must also be local (even if they locally violate a lemma which is not yet explicit in the proof).

Lakatos's identification of counterexamples which are global but not local shows that it is, in some way, possible in mathematics to rebut without a known undercutter (see ([1, p. 298] and [38, pp. 22–24] for previous discussion on this). However, his recommendation for dealing with such counterexamples suggests that he thought that it was possible to see such entities as undercutters as well as rebutters, with their role being to make a misleading proof more precise. It is an interesting question as to whether this situation, in which any rebutter *must be* an undercutter, either explicitly or implicitly, is specific to mathematics. Certainly, in other areas of thought, there are many examples which appear to rebut without undercutting. The British journalist Bruce Anderson provides a rather nice example of this:

---

[14] An additional clue that the cylinder may have been conceived by Lakatos is that when GAMMA first appears [23, p. 8] it is to raise questions about Cauchy's proof of the Euler conjecture, to which Lakatos appends the following footnote: "The class is a rather advanced one. To Cauchy, Poinsot, and to many other excellent mathematicians in the nineteenth century these questions did not occur." Of course, it doesn't follow that *all* of GAMMA's interjections are anachronistic, but it is interesting that he's introduced in this way.

> A generation ago, Reg Prentice was minister of education in a Labour government. He wrote a paper making the case for increased spending on nursery schools. The Chief Secretary was Joel Barnett, a tough little Mancunian accountant. 'Reg,' he said, 'brilliant paper. It was so moving; reminded me why I came into the Labour movement. Your arguments are unanswerable. And the answer's no.' [4]

We contend that Barnett's response would be untenable in mathematics (but perhaps only mathematics).

Our analysis of Lakatos's proof-changing methods from a Waltonian perspective might resolve the rebutting-without-undercutting problem (see §13 below).

13.4. Lakatos's method of proofs and refutations. Lakatos's method of *proofs and refutations* extends his *lemma-incorporation*, in which he suggests using the proof steps to find counterexamples (by looking for objects which would violate them). For any counterexamples found, one should determine whether they are local or global counterexamples, and then perform lemma-incorporation. This may correspond to using Walton's defeating questions to suggest premisses which may be mistaken or incomplete.

13.5. Extending Walton's schemes. In sections 13.5.3 to 13.5.5 we suggest extensions to Walton's schemes, which reflect Lakatos's proof-changing methods. In order to understand them, it is useful to suggest a new procedural scheme first (§13.5.1). We can then instantiate this to Cauchy's proof plan (§13.5.2). The procedural proof outlined by Cauchy may seem a rather curious beast, in that it starts off with one type of object (a polyhedron), performs various operations on the object (which change it from a polyhedron to a network, to a triangulated graph), shows that the resulting object, now of a different type, has a certain property, and concludes from this that the original object must have another property (Eulerianness). However, this is a common proof strategy in mathematics: arguably, this is *the* signature proof strategy of contemporary mathematics, perhaps beginning with Galois theory. Marquis [34] discusses this general approach, and gives a worked example considering K-theory:

> In algebraic K-theory, one starts with the category of rings and ring homomorphisms, then associates to each ring a commutative semigroup

> and to each ring homomorphism a semigroup homomorphism, then applies the above functor to end up in the category of abelian groups and group homomorphisms, and thus obtains information about some of the structural properties of these rings. [34, p. 257]

Another example of this type of reasoning is the application of Galois theory to demonstrate the impossibility of constructions in Euclidean geometry. Indeed, Galois's development of such techniques is arguably a critical turning point in mathematical method.

We can see these as examples of analogical reasoning in mathematics, in which there is a known, well-defined, bijective mapping between a source and target domain. This is a specific and important type of analogy: inferences made in a target domain can be used to infer certain knowledge about the source domain. An example of a known, well-defined, bijective mapping is a symmetrical transformation such as scaling, reflection, and rotation. For instance, analogical inferences about a kite-shaped quadrilateral with its apex pointing upwards and a kite-shaped quadrilateral with its apex pointing right, where the mapping between the two shapes is a rotation of 90°, will be rigourous. While analogical reasoning is ubiquitous, mathematics is probably a unique domain in that there exist known, well-defined, bijective mappings, and therefore analogical inference in this context is a rigourous form of reasoning: in other domains there is often a 'verification' stage after the analogical inference (clearly this can also occur in mathematical domains in which the mappings are not well-defined). As such, we can see our following scheme, *Suggested Argumentation Scheme for a Procedural Argument*, as a subcategory of Walton's *Argument from Analogy* [54, p. 315] mentioned in §9. Our scheme differs from Walton's schemes in a few regards. Most importantly, it is procedural. Secondly, the scheme contains multiple inferences: Walton's schemes in [54, Ch. 9] describe just one inference (although some of his examples in [54, Ch. 10] describe multiple inferences, such as *Argument from Guilt by Association*, in which premise 2 follows from premise 1). Thirdly, because of the second point, the order of the premisses matters. One possibility to bring our scheme closer to a Waltonian scheme would be to split it into several schemes: we present it as a single scheme for ease of understanding.

13.5.1. Suggested Argumentation Scheme for a Procedural Argument.
Premise 1 Take an arbitrary x such that Px, say m. Then:
  (i) $f_1 : P \to T_1$, $f_1(m) = m_1$ and $P_1 m_1$,

(ii) $f_2 : T_1 \to T_2$, $f_2(m_1) = m_2$ and $P_2 m_2$,
(iii) $f_3 : T_2 \to T_3$, $f_3(m_2) = m_3$ and $P_3 m_3$,
$\vdots$
(n) $f_n : T_{n-1} \to T_n$ and $f_n(m_{n-1}) = m_n$ and $P_n m_n$, for functions f_1 to f_n, types T_1 to T_n and properties P_1 to P_n.

Premise 2 Therefore (since m was arbitrary), there exist functions f_1 to f_n st $\forall x(Px \to \exists y \text{ st } (f_n(f_{n-1}(\ldots f_2(f_1(x)))))) = y) \wedge P_n y)$.

Premise 3 $(P_n m_n \to P_{n-1}(f_n^{-1}(m_n))) \wedge (P_{n-1}(m_{n-1})) \to P_{n-2}(f_{n-1}^{-1}(m_{n-1}))) \wedge \cdots \wedge (P_2 m_2 \to P_1(f_2^{-1}(m_2))) \wedge (P_1 m_1 \to E(f_1^{-1}(m_1)))$.

Conclusion $\forall x(Px \to Ex)$.

Critical Questions:

1. Is m really arbitrary? Does it have any properties that other objects with property P might not have?
2. Is the function f_i well-defined, for all i? Does it always take objects of type T_i and output objects of type T_{i+1}?
3. Does property P_n always hold for m_n?
4. What is the relationship between P_n and E? Can we conclude that because Pm and $P_n m_n$, that Em?

We instantiate this scheme in terms of Lakatos's description of Cauchy's proof sketch, rephrased from [23, pp. 7–8], as the following procedural scheme, with critical questions all taken from [23, p. 8] (clearly there would be further critical questions too). Note that the third function in step (iii) is applied multiple times.

13.5.2. Example Instantiation of the Argumentation Scheme for Procedural Argument.

Premise 1 Take the cube, which is an arbitrary polyhedron. Then:
(i) Remove a face from the cube. We can stretch the remaining surface flat on the blackboard, and the Euler characteristic of this surface will be that of the cube, minus one (as we removed a face).
(ii) Triangulate the connected network from step (i). The Euler characteristic from step (i) is preserved (since for any new edge which is added, one will always get a new face).
(iii) Drop the triangles one by one from the triangulated map from step (ii). Again, the Euler characteristic from step (ii) is preserved (since there are only two alternatives — the disappearance of one edge

and one face, or else of two edges, one vertex and a face). Apply this step repeatedly (n times) until a single triangle remains.

Premise 2 The triangle from step (*iii*) has an Euler characteristic of one (since $V - E + F = 3 - 3 + 1 = 1$).

Premise 3 Therefore for any polyhedron, there exist procedures (*i*), (*ii*) and (*iii*) which will turn the polyhedron into a single triangle, and that triangle has an Euler characteristic of one.

Premise 4 If we start with a triangle, which has an Euler characteristic of one, then there is a way in which we can add triangles one by one (n times), resulting in a triangulated map each time and preserving the Euler characteristic (we either add one edge and one face, or else of two edges, one vertex and a face). There is a way of removing edges to "untriangulate" the map. The Euler characteristic is preserved from the previous step (since for any new edge which is removed, one will always get one less face). There is now a way in which we can add a face and assemble the network into a polyhedron. Since there is one more face, the Euler characteristic is now two.

Conclusion All polyhedra have an Euler characteristic of two.

Critical Questions:

1. Does the cube have particular properties which allow us to perform this procedure, which other polyhedra do not share? For instance, can we perform step (*i*) on any polyhedron, with the property that the resulting surface can be stretched flat on a board? (Proposed by ALPHA [*ibid.*])
2. When we perform step (*ii*), does the property of preserving the Euler characteristic always hold (do we always get a new face for any new edge)? (Proposed by BETA [*ibid.*])
3. When we perform step (*iii*), does the property of preserving the Euler characteristic always hold (do we always remove one edge and one face, or two edges, one vertex and a face)? Are there always a countable number of repetitions of step (*iii*) which will result in a single triangle? (Proposed by GAMMA [*ibid.*])

We can now outline our suggested argumentation schemes for Lakatos's proof-changing methods. In all three schemes, initially only the first two premisses are noted: given these, Lakatos suggests doing the work involved to find the third premise and then the conclusion may be reached.

13.5.3. Suggested Argumentation Scheme for Global and Local Lemma-Incorporation.

Premise C_1 is a global counterexample to conjecture P, that is $C_1 \Rightarrow \neg P$.

Premise C_1 is a local counterexample to conjecture P, that is at least one of the premisses on which P depends, $Prem_i$ is itself a conclusion in an (instantiation of) a scheme which contains critical question(s) which must receive a defeating answer if C_1 holds.

Premise X is the concept of *objects for which* $Prem_i$ *holds.*

Conclusion Replace P by $X \Rightarrow P$.

Note that if P is the proposition $R \Rightarrow Q$, then the replacement is $(R \wedge X) \Rightarrow Q$, which is clearly in line with our Lakatosian example. As noted earlier (p. 43) this is a form of strategic withdrawal in which the concept to which one withdraws is provided by the notion of a problematic premise being satisfied.

13.5.4. Suggested Argumentation Scheme for Local and not Global Lemma-Incorporation.

Premise C_1 is a local counterexample to conjecture P, that is at least one of the (instantiations of) the schemes upon which the derivation of P depends, $Prem_i$, contains critical question(s) CQ_i which must receive a defeating answer if C_1 holds.

Premise C_1 is not a global counterexample to conjecture P, that is $C_1 \not\Rightarrow \neg P$.

Premise X is the concept of *objects for which* $Prem_i$ *fails*

Conclusion Replace $Prem_i$ by $\neg X \Rightarrow Prem_i$.

Again, note that if $Prem_i$ is the proposition $R \Rightarrow S$, then the replacement is $(R \wedge \neg X) \Rightarrow S$, which is in line with the Lakatosian example. We also noted earlier (p. 44) that this is a form of piecemeal exclusion, in which the concept which one excludes is designed to exactly cover the cases in which a problematic premise fails.

13.5.5. Suggested Argumentation Scheme for Global and not Local Lemma-Incorporation.

Premise C_1 is a global counterexample to conjecture P, that is $C_1 \Rightarrow \neg P$.

Premise C_1 is not a local counterexample to conjecture P, that is none of the (instantiations of) schemes upon which the derivation of P depends contains critical question(s) CQ_i which must receive a defeating answer if C_1 holds ($\forall i$ st $1 \leqslant i \leqslant n, P_i \Rightarrow C_1$, where n is the number of premisses).

Premise $Prem_i$ is a premise in the derivation of P which contains an unstated assumption, A_1 such that $(A_1 \wedge Prem_i) \Rightarrow \neg C_1$.

Conclusion Replace P by $A_1 \Rightarrow P$, and $Prem_i$ by $A_1 \wedge Prem_i$.

The absence of an undercutter follows from the incompleteness of the set of critical questions. Hence, in a narrow sense, we can have rebutting without undercutting. But the incompleteness can always (in principle) be remedied: there is always an A_1 for the proof analysis to find. Once it has been remedied, the undercutter has been found, since $C_1 \Rightarrow \neg A_1$ (by transposition), and thus the question 'Does A_1 hold?' must receive a defeating answer if C_1 holds.

The unstated assumption may well be a further property in one of the proof steps (from premise 1 in argumentation scheme in §13): this is the case in Lakatos's examples. That is, $f_i \colon T_{i-1} \to T_i$ and $f_n(m_{i-1}) = m_i$ and $P_i m_i$ *and* there is a further property P' such that $P' m_i$, for some $1 \leqslant i \leqslant n$. Initially this further property is not explicit, but it holds for all (known) positive examples and fails for all (known) global but not local counterexamples. The patch is then to make it explicit, by adding it to the relevant proof step. The presence of an unstated assumption in the derivation suggests that the set of critical questions is incomplete. (At some point the question 'Does A_1 hold?' should have been posed, since a negative answer would undercut the derivation.) There is nothing intrinsically wrong with this — it is *informal* logic after all — but it does indicate that the proof analysis is shoddy, which is of course Lakatos's point.

Closing remarks

Whether we are schemifying Toulmin or Toulminizing abduction, an analysis of where our five protagonists may agree and disagree suggests all kinds of fascinating new areas. Connections between the different theories have typically been neglected, to the extent that our most contemporary thinker, Walton, is the only author to discuss any of the others in the works we cite, and even he does not consider Lakatos. Finding and expanding such connections may well be a fruitful avenue of research, so that different descriptions of our elephant may, possibly, begin to converge. Additionally, we hope that our particular breed of interest, the mathematical elephant, will prove to be similar in many areas to other breeds, and both informal logic and the philosophy of mathematical practice will benefit. In any case, we shall enjoy enormously the attempt to describe it in such a way.

Acknowledgments. We are grateful to the Mathematical Reasoning Group at the University of Edinburgh for a lively debate on some of the ideas presented here. In particular, we are grateful to Alan Smaill and Markus Guhe for their thoughts and comments on earlier drafts of this work. We would also like to thank the organisers of the conference on *Argumentation as a cognitive process*, held at Nicolaus Copernicus University, Toruń, Poland on 13–15 May 2010, for all their work in bringing such an interesting community together. Alison Pease is supported by EPSRC grant EP/F035594/1.

References

[1] Aberdein, A., "The uses of argument in mathematics", *Argumentation* 19 (2005): 287–301.

[2] Aberdein, A., "Managing informal mathematical knowledge: Techniques from informal logic", pages 208–221, in: J. M. Borwein and W. M. Farmer (eds.), *MKM 2006, LNAI 4108*, Springer-Verlag, Berlin, 2006.

[3] Banegas, J. A., "L'argumentació en matemàtiques", pages 135–147 in: E. Casaban i Moya (ed.), *XIIè Congrés Valencià de Filosofia*, València, 1998. http://my.fit.edu/~aberdein/Alcolea.pdf

[4] Anderson, B., "Diary", *Spectator* 313 (9481), May 15th 2010. http://www.spectator.co.uk/politics/all/5996573/part_2/diary.thtml

[5] Bocheński, I. M., *A History of Formal Logic*, Chelsea Pub Co, New York, N.Y., 1970 [1956].

[6] Boden, M. A., *The Creative Mind: Myths and Mechanisms*, Weidenfield and Nicholson, London, 1990.

[7] Burley, W., *On the Purity of the Art of Logic: The Shorter and Longer Treatises*, Yale University Press, New Haven, CT, 2000. Translated by P. V. Spade.

[8] Cauchy, A. L., *Cours d'Analyse de l'École Royale Polytechnique*, de Bure, Paris, 1821.

[9] Colton, S., *Automated Theory Formation in Pure Mathematics*, Springer-Verlag, 2002.

[10] Colton, S., A. Bundy, and T. Walsh, "On the notion of interestingness in automated mathematical discovery", *International Journal of Human Computer Studies* 53, 3 (2000): 351–375.

[11] Conan Doyle, A., *The Memoirs of Sherlock Holmes*, Forgotten Books, 2008 [1894].

[12] Dauben, J. W., "Peirce's place in mathematics", *Historia Mathematica* 9, 3 (1982): 311–325.

[13] Dove, I. J., "On mathematical proofs and arguments: Johnson and Lakatos", pages 346–351 in: F. H. Van Eemeren and B. Garssen (eds.), *Proceedings of the Sixth Conference of the International Society for the Study of Argumentation*, volume 1, Sic Sat, Amsterdam, 2007.

[14] Dunham, W., *Euler: The Master of Us All*, The Mathematical Association of America, Washington, DC, 1999.

[15] Epstein, S. L., "Learning and discovery: One system's search for mathematical knowledge", *Computational Intelligence* 4, 1 (1988): 42–53.

[16] Gasteren, A. J. M., *On the shape of mathematical arguments*, Lecture notes in computer science, volume 445, Springer, Berlin, 1990.

[17] Gray, J., *The Hilbert Challenge*, Oxford University Press, Oxford, 2000.

[18] Grosholz, E. R., *Representation and Productive Ambiguity in Mathematics and the Sciences*, Oxford University Press, New York, NY, 2007.

[19] Hairer, E., and G. Wanner, *Analysis by Its History*, Undergraduate Texts in Mathematics, Springer, New York, NY, 2008.

[20] Hanson, N. R., "Is there a logic of scientific discovery?", pages 20–35 in: H. Feigl and G. Maxwell (eds.), *Current Issues in the Philosophy of Science*, Symposia of scientists and philosophers. Proceedings of Section L of the American Association for the Advancement of Science, 1959, Holt, Rinehart and Winston, New York, NY, 1961.

[21] Heath, T. L., *The Thirteen Books of Euclid's Elements*, Cambridge University Press, Cambridge, 2nd edition, 1925.

[22] Hitchcock, D., "Toulmin's warrants", pages 69–82 in: F. H. van Eemeren, J. Blair, C. Willard, and A. F. Snoeck-Henkemans (eds.), *Anyone Who Has a View. Theoretical Contributions to the Study of Argumentation*, Kluwer Academic Publishers, Dordrecht, 2003.

[23] Lakatos, I., *Proofs and Refutations*, Cambridge University Press, Cambridge, 1976.

[24] Lakatos, I., "Cauchy and the continuum: The significance of non-standard analysis for the history and philosophy of mathematics", pages 43–60 in: J. Worral and C. Currie (eds.), *Mathematics, science and epistemology*, Philosophical Papers, volume 2, Cambridge University Press, Cambridge, 1978.

[25] Lakatos, I., "Proofs and refutations (I)", *The British Journal for the Philosophy of Science* 14 (1963): 1–25.

[26] Lakatos, I., "Proofs and refutations (II)", *The British Journal for the Philosophy of Science* 14 (1963): 120–139.

[27] Lakatos, I., "Proofs and refutations (III)", *The British Journal for the Philosophy of Science* 14 (1963): 221–245.
[28] Lakatos, I., "Proofs and refutations (IV)", *The British Journal for the Philosophy of Science* 14 (1964): 296–342.
[29] Larvor, B., *Lakatos: An Introduction*, Routledge, London, 1998.
[30] Lenat, D., *AM: An Artificial Intelligence Approach to Discovery in Mathematics*, PhD thesis, Stanford University, 1976.
[31] Leng, M., "What's there to know? A fictionalist account of mathematical knowledge", pages 84–108 in: M. Leng, A. Paseau, and M. Potter (eds.), *Mathematical Knowledge*, Oxford University Press, Oxford, 2007.
[32] Macedo L., and A. Cardoso, "Creativity and surprise", in: G. Wiggins (ed.), *Proceedings of the AISB'01 Symposium on Artificial Intelligence and Creativity in Arts and Science*, 2001.
[33] Mancosu, P., "Mathematical explanation: Problems and prospects", *Topoi* 20, 1 (2001): 97–117.
[34] Marquis, J.-P., "Abstract mathematical tools and machines for mathematics", *Philosophia Mathematica* 5 (1997): 250–272.
[35] Muntersbjorn, M., "Construction, articulation, and explanation: Phases in the growth of mathematics", pages 19–42 in: B. Van Kerkhove, J.-P. van Bendegem, and J. De Vuyst (eds.), *Philosophical Perspectives on Mathematical Practice*, College Publications, London, 2010.
[36] Pease, A., *A Computational Model of Lakatos-style Reasoning*, PhD thesis, School of Informatics, University of Edinburgh, 2007. http://hdl.handle.net/1842/2113
[37] Pease, A., M. Guhe, and A. Smaill, "Analogy formulation and modification in geometry", pages 358–364 in: *Proceedings of the Second International Conference on Analogy*, 2009.
[38] Pease, A., A. Smaill, S. Colton, and J. Lee, "Bridging the gap between argumentation theory and the philosophy of mathematics", *Foundations of Science* 14, 1–2 (2009): 111–135.
[39] Peirce, C. S., *Collected Papers of Charles Sanders Peirce*, Eight Volumes, Harvard University Press, Cambridge, Mass, 1931–58.
[40] Peirce, C. S., *The New Elements of Mathematics*, Mouton, The Hague, 1976.
[41] Plato, *The Republic*, OUP, Oxford, 1993.
[42] Pollock, J., "The structure of epistemic justification", *American Philosophical Quarterly, monograph series* 4 (1970): 62–78.
[43] Pollock, J., *Contemporary Theories of Knowledge*, Rowman and Littlefield, Totowa, NJ, 1986.
[44] Pollock, J., *Cognitive Carpentry*, The MIT Press, Cambridge, MA., 1995.
[45] Polya, G., *Mathematics and plausible reasoning*, volume 1, Induction and analogy in mathematics, Princeton University Press, 1954.

[46] Quine, W. V. O., "Two dogmas of empiricism", *From a Logical Point of View*, Harvard University Press, 1953. First version of the paper, without any reference to Duhem, in *The Philosophical Review* 60 (1951): 20–53.

[47] Reed, C., and G. Rowe, "Translating Toulmin diagrams: Theory neutrality in argument representation", *Argumentation* 19, 3 (2005): 267–286.

[48] Robinson, A., *Non-standard Analysis*, Princeton University Press, Princeton, New Jersey, 1996. Revised Edition.

[49] Robinson, R., "Analysis in Greek geometry", *Mind* 45 (1936): 464–73.

[50] Sandifer, C. E., *The early mathematics of Leonhard Euler*, The Mathematical Association of America, 2007.

[51] Seidel, P. L., "Note über eine Eigenschaft der Reihen, welche Discontinuirliche Functionen Darstellen", *Abhandlungen der Mathmatisch-Physikalischen Klasse der Königlich Bayerischen Akademie der Wissenschaften* 5 (1847): 381–93.

[52] Toulmin, S., *The uses of argument*, CUP, Cambridge, 1958.

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

[54] Walton, D., C. Reed, and F. Macagno, *Argumentation Schemes*, Cambridge University Press, Cambridge, 2008.

[55] Walton, D. N., *Argument Schemes for Presumptive Reasoning*, Lawrence Erlbaum Associates, Mahwah, NJ, 1996.

[56] Walton D. N., and C. A. Reed, "Diagramming, argumentation schemes and critical questions", pages 195–211 in: F. H. van Eemeren, J. Anthony Blair, Ch. A. Willard, and A. Francisca Snoeck Henkemans (eds.), *Anyone Who Has a View: Theoretical Contributions to the Study of Argumentation*, Kluwer Academic Publishers, Dordrecht, 2003.

ALISON PEASE
Centre for Intelligent Systems and their Applications
Informatics Forum
University of Edinburgh
8 Crichton Street
Edinburgh, EH8 9AB
A.Pease@ed.ac.uk

ANDREW ABERDEIN
Department of Humanities and Communication
Florida Institute of Technology
150 West University Blvd
Melbourne, Florida 32901-6975, U.S.A.
aberdein@fit.edu