# Lakatos-style Reasoning

Alison Pease, Simon Colton, Alan Smaill, John Lee
Division of Informatics, University of Edinburgh,
80 South Bridge, Edinburgh, EH1 1HN,UK
{alisonp@dai., simonco@dai.,A.Smaill@,john@cogsci.}ed.ac.uk

A problem in modelling mathematics is that few people have analysed and reported the way in which mathematicians work. Lakatos(4) is a welcome exception. He presents a rational reconstruction of the evolution of Euler's conjecture that for all polyhedra, the number of vertices ($V$) minus the number of edges ($E$) plus the number of faces ($F$) is two, and its proof. This work spans 200 years of concepts, conjectures, counter-examples and 'proofs' and is invaluable to AI researchers trying to model mathematical reasoning. Such models might serve to *(a)* illuminate aspects of human mathematics, and *(b)* improve existing automated reasoning programs.

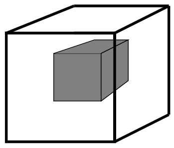

Figure 1: The hollow cube; $V - E + F$ = 16 - 24 + 12 = 4

Lakatos suggests that Euler's conjecture may have been proposed after noticing that the relationship $V - E + F = 2$ holds for all regular polyhedra (the tetrahedron, cube, octahedron, icosahedron and dodecahedron). The counter-example of the hollow cube - a cube with a cube-shaped hole in it - is then found. Reactions to this include:

1) saying that the existence of the counter-example is sufficient to disprove the conjecture, and therefore abandon it;

2) claiming that the hollow cube is *not* a counter-example since it is not a polyhedron. The conjecture remains unchallenged and the definition of polyhedron is tightened from 'a solid whose surface consists of polygonal faces' to 'a surface consisting of a system of polygons' (thus excluding the hollow cube);

3) generalising from the hollow cube to 'any polyhedra with a cavity' and then excluding these from the conjecture. The new conjecture is 'for all polyhedra *without cavities*, $V - E + F = 2$';

4) generalising from regular polyhedra to 'convex polyhedra' and then limiting the conjecture to these. The new conjecture is 'for all *convex* polyhedra, $V - E + F = 2$'.

Lakatos categorises these responses into methods: *induction* (scientific rather than mathematical) for generating the initial conjecture; and then if a counter-example is found: 1) *surrender*; 2) *monster-barring*; and two types of *exception-barring* - 3) *piecemeal exclusion* and 4)*strategic withdrawal*. Our aim is to model these methods, using the theory formation program HR(2) as a base. HR starts with objects of interest (*e.g.*, integers) and initial concepts (*e.g.*, multiplication and addition) and uses production rules to transform old concepts into new ones. For example a production rule might take the concept 'number of divisors of an integer', and change it to the concept 'number of divisors of an integer = 2'. It would then list all integers which share this property (*i.e.*, all the primes in its database). Conjectures, such as concept $X$ implies concept $Y$, are made empirically by comparing the example sets of different concepts. For instance HR has made the conjecture that if the sum of the divisors of $n$ is prime, then the number of divisors of $n$ is prime (2). Since HR works especially well in number theory we have applied the methods to this domain. *Induction* can be used to generate conjectures of the form $\forall x$, $P(x)$; and $-\exists x$ such that $P(x)$, for some property $P$. Fermat's Last Theorem is an example of this kind of conjecture. Clearly *surrender* is sometimes necessary (many of the conjectures and concepts generated will not be worth modifying) although few surrendered conjectures are recorded. One example in (1) is the conjecture that the *nth* perfect number $P_n$ contains $n$ digits. This was found by induction on the first four perfect numbers - 6, 28, 496, 8128 - but surrendered when the fifth - $33,550,336$ -

was found. An example of *monster-barring* is in (3), in which the definition of a prime number is modified from 'a natural number that is only divisible by itself and 1' to 'a natural number with exactly two divisors'. This occurs because the number 1 (considered prime in the first but not in the second definition) is a counter-example to many conjectures. For example in the Fundamental Theorem of Arithmetic, that every natural number is either prime or can be expressed uniquely as a product of primes, the uniqueness criterion is violated ($6 = 2 \times 3 = 2 \times 3 \times 1 = 2 \times 3 \times 1 \times 1$ etc). Rather than explicitly exclude 1 in the theorem, it is more elegant to exclude it from the concept definition, and thus the current definition of a prime is formed and accepted. Examples of *exception-barring* are plentiful, for instance Goldbach's conjecture that all even numbers *except 2* can be expressed as the sum of two primes; all primes *except 2* are odd; and all integers *except squares* have an even number of divisors.

So far we have modelled simple versions of both types of exception-barring in HR. To do this we have had to enable it to generate conjectures with known counter-examples (whereas previously only conjectures true of *all* examples were made). We have done this in two ways (and are currently experimenting to see which is preferable). Firstly we have enabled it to make *near equivalence* conjectures, *i.e.*, conjectures which hold for $x\%$ of the objects of interest (where $x$ is defined by the user). Secondly we have run two versions of HR, which are able to communicate concepts and conjectures to each other, in parallel. These simple agents have access to different databases and therefore one may make a conjecture which is true of all its examples and ask the other for counter-examples. If there is a large number of counter-examples then HR will look to see if it already has a concept which includes only the counter-examples, and then use *piecemeal exclusion*. For instance with a database of numbers $1-30$ it formed the near equivalence conjecture that 'all integers have an even number of divisors', with counter-examples 1, 4, 9, 16 and 25. It then found the concept of 'square numbers' which covers the counter-examples and modified the conjecture to 'all integers except squares have an even number of divisors', which is true. If HR makes a conjecture for which there are only a few counter-examples (where the number is specified by the user), it will invent a concept with a definition which excludes the counter-examples and then perform *strategic withdrawal*. When we ran the agent version with databases $1-20$ and $20-40$ the second agent generated the conjecture 'all primes numbers are odd' and sent it to the first, which found the counter-example 2. The second agent then generated the new concept 'primes numbers except 2' and modified the conjecture to 'all primes numbers except 2 are odd'.

These preliminary results look promising. Although other systems can be said to use certain methods (HR(2) already used *induction* and Lenat's AM program(5) performed a kind of *monster-barring*), this is the only dedicated attempt we know of to model Lakatos-style reasoning.

## References


[1] D. Burton. *The History of Mathematics.* Allyn and Bacon, Inc, Boston, USA, 1985.

[2] S. Colton. *Automated Theory Formation in Pure Mathematics.* PhD thesis, Dept. of Artificial Intelligence, University of Edinburgh, 2001.

[3] C. Dunmore. Meta-level revolutions in mathematics. In D. Gillies, editor, *Revolutions in Mathematics*, pages 209–225. Clarendon Press, Oxford, 1992.

[4] I. Lakatos. *Proofs and Refutations.* CUP, 1976.

[5] D. Lenat. *AM: An Artificial Intelligence Approach to Discovery in Mathematics.* PhD thesis, Stanford University, 1976.


## Acknowledgements


This work was supported by EPSRC grants GR/M45030. and GR/M98012. The second author is also affiliated with the Department of Computer Science, University of York.