Seventy four minutes of mathematics: An analysis of the third Mini-Polymath project

Alison Pease and Ursula Martin

{Alison.Pease, Ursula.Martin}@eecs.qmul.ac.uk
School of Electronic Engineering and Computer Science,
Queen Mary, University of London
Mile End Road London E1 4NS
UK

Abstract. Alan Turing proposed to consider the question, "Can machines think?" in his famous article [40]. We consider the question, "Can machines do mathematics, and how?" Turing suggested that intelligence be tested by comparing computer behaviour to human behaviour in an online discussion. We hold that this approach could be useful for assessing computational logic systems which, despite having produced formal proofs of the Four Colour Theorem, the Robbins Conjecture and the Kepler Conjecture, have not achieved widespread take up by mathematicians. It has been suggested that this is because computer proofs are perceived as ungainly, brute-force searches which lack elegance, beauty or mathematical insight. One response to this is to build such systems which perform in a more human-like manner, which raises the question of what a "human-like manner" may be.

Timothy Gowers recently initiated Polymath [4], a series of experiments in online collaborative mathematics, in which problems are posted online, and an open invitation issued for people to try to solve them collaboratively, documenting every step of the ensuing discussion. The resulting record provides an unusual example of fully documented mathematical activity leading to a proof, in contrast to typical research papers which record proofs, but not how they were obtained.

We consider the third Mini-Polymath project [3], started by Terence Tao and published online on July 19, 2011. We examine the resulting discussion from the perspective: what would it take for a machine to contribute, in a human-like manner, to this online discussion? We present an account of the mathematical reasoning behind the online collaboration, which involved about 150 informal mathematical comments and led to a proof of the result. We distinguish four types of comment, which focus on mathematical concepts, examples, conjectures and proof strategies, and further categorise ways in which each aspect developed. Where relevant, we relate the discussion to theories of mathematical practice, such as that described by Pólya [36] and Lakatos [24], and consider how their theories stand up in the light of this documented record of informal mathematical collaboration.

Keywords: Turing test, mathematical practice, polymath, mathematics, philosophy of mathematics, automated theorem proving, automated theory formation, computational logic systems.

1 Collaborative mathematics and Turing

From 8pm to 9.14pm on July 19th 2011, twenty seven people from around the world took part in an online experiment: could they collaboratively solve a problem of International Mathematical Olympiad (IMO) standard? The answer was a resounding yes, and discussion of both the problem and the novel way of solving it continued long after a proof was discovered. This experiment was called Mini-Polymath 3 and is part of a series of experiments led by Timothy Gowers and Terence Tao on massive online collaboration in mathematics.

In this paper we consider the question: what would it take for a computational agent to contribute, in a human-like manner, to a Mini-Polymath discussion? That is, we imagine a machine which were able to translate a mathematical problem into another domain, spot a typo in the original problem and suggest a correction, explain to another participant why a particular object is a counterexample to a conjecture, express itself at an appropriate level of detail and expand if necessary, find a proof strategy and compare it to an everyday object, appreciate a promising direction outlined by someone else, and so on. These (and many more) remarkable capabilities shown by participants in Mini-Polymath 3 raise the question of what sort of functionality automated mathematicians would need to develop, in order to participate in such a dialogue. Turing [40] suggested that one way of measuring whether a machine is intelligent is by comparing its behaviour to human behaviour. We explore his proposal in [35]: our starting point in this paper is that a machine which were able to do some of the remarkable things which humans do while engaged in mathematical collaboration would be an enormous advance on current theorem-proving technology. This leads to the main contribution of our paper: an analysis of the discussion which took place over the third Mini-Polymath project.

2 The Polymath Projects

In 2009 the mathematician Timothy Gowers posted the question "Is massively collaborative mathematics possible?" on his blog [19]. He and Terence Tao then initiated a series of experiments on collaborative mathematics by posting open, difficult conjectures and inviting readers to collaboratively prove them. Gowers asked that people follow a set of guidelines (which had themselves emerged as the result of an online collaborative discussion). These were designed to encourage massively collaborative mathematics both in the sense of involving as many people as possible: "we welcome all visitors, regardless of mathematical level, to contribute to active polymath projects by commenting on the threads" [1]; and having a high degree of interaction, with results arising from the rapid exchange of informal ideas, as opposed to parallelisation of sub-tasks: rules 3 and 6 state that "It's OK for a mathematical thought to be tentative, incomplete, or even incorrect." and "An ideal polymath research comment should represent a 'quantum of progress'." (*ibid.*).

These experiments have been successful in that they have sometimes led to a proof of the conjecture or, in the case of the first Polymath experiment, to a proof of a more general conjecture. The type of collaboration was also along the lines that Gowers had outlined and the project developed much more quickly than a standard proof attempt. This seems to be partly because it gathered momentum (some spoke of the excitement of the project progressing while they slept) and partly because comments from other authors led to contributors having thoughts they otherwise might not have had. For instance, Gowers gives an example of one contributor developing ideas in a domain which he was not familiar with (ergodic theory), and another who translated these ideas into one that he was familiar with (combinatorics), thus affecting his own line of reasoning. However, Gowers expressed some disappointment at the level of collaboration: in [18] he said that despite wide participation initially, the number of contributors settled down to a handful, all of whom he knew personally (Nielsen [33] hypothesises that the reason for this is that academics do not have time to contribute to blogs, since this type of contribution is not recognised by the rewards system). There have been six Polymath projects to date, and three publications (all describing work carried out in the first project). These blogs and the philosophy which underlies them have enabled a level of collaboration which, before the internet, would probably have been impossible to achieve; the open invitation has widened the mathematical community, as amateurs can now participate in mathematical discussions with experts; and the focus on short informal comments has resulted in a readily available and public record of mathematical progress. As Gowers said, it is possibly "the first fully documented account of how a serious research problem was solved, complete with false starts, dead ends etc." [18]. The answer to Gowers' question then, is "yes (although conditions which facilitate it can be improved)".

The Mini-Polymath projects were a spin off from the Polymath research projects, to explore whether the same sort of collaboration could be effective in an educational setting. Tao posted an IMO question on his blog: participation was open to anyone, and participants were asked to follow the same guidelines as Gowers had set out. The fact that the problem was posed in an educational context results in a very different dynamic to a research context. Participants (students) are committed to proving the conjecture: they trust it to be provable, and of an appropriate standard.

There have been three Mini-Polymath projects to date: all solved the problem. In sections 3 to 9 we analyse the discussion in the third Mini-Polymath. In section 10 we discuss two theories of mathematical practice, by Pólya [36] and Lakatos [24], and consider whether their theories are a good description of the Mini-Polymath collaboration. Finally, in section 11, we consider what functionality a computational blogger would need in order to participate in the discussion.

3 Mini-Polymath 3

The problem below was composed by Geoff Smith, specifically for the 52nd IMO which was held in Amsterdam, Netherlands, in July 1324, 2011. It formed question 2 of the IMO, given to contestents on Monday July 18th. Tao posted the problem at 8pm on July 19th, 2011 Coordinated Universal Time (UTC)[1], having posted in advance that he would do so. The relevant websites are the *research thread* [3] hosted at the polymath blog for the problem solving process, a *discussion thread* [2] hosted at Terence Tao's blog for any meta-discussion about the project, and a *wiki page* [5] hosted by Michael Nielsen for a summary of the problem and discussion.

> Let S be a finite set of at least two points in the plane. Assume that no three points of S are collinear. A *windmill* is a process that starts with a line l going through a single point $P \in S$. The line rotates clockwise about the pivot P until the first time that the line meets some other point Q belonging to S. This point Q takes over as the new pivot, and the line now rotates clockwise about Q, until it next meets a point of S. This process continues indefinitely.
> Show that we can choose a point P in S and a line l going through P such that the resulting windmill uses each point of S as a pivot infinitely many times. (Tao, 8:00pm)

Interest was immediate and progress was quick: seventy four minutes after posting it the participants had found a solution. For the most part, people followed the rules; these were largely self regulating anyway, due to the speed of responses: long answers in which someone had spent too long working individually were often ignored simply because they appeared after everyone else had moved on. Tao wrote the 27th comment, in which he recognised an argument given in the 14th comment as a proof. This argument was posted at 9.14pm, and Tao posted his "official" recognition of its status as proof at 9.50pm. Although there were subsequent comments, we focus here on comments posted between 8 and 9.14, and 9.14 and 9.50. Only comments which start a thread are numbered in the blog: we follow that numbering but introduce letters to identify comments within a thread.

4 A typology of comments

Between the time at which Tao posted the problem (8pm) and the time he posted that it had been solved (9.50pm), there were 147 comments over 27 threads. We categorised each comment according to whether it mainly concerned a concept, example, conjecture or proof. For instance, in the comments shown below, 18a is concerned the invention of a new concept, 3a with an example, 11a and 11c with the conjecture and 4a with the proof. This is a rather subjective categorisation: while some comments (such as those shown below) seem straightforward, others required more interpretation. Ten comments fell into more than one category, for instance we categorised 23a as both forming an (implicit) conjecture and as concerning examples. Figure 1 shows the proportion of each category.

[1] All timepoints given hereafter refer to this date.

18a. Since the points are in general position, you could define the wheel of p, w(p) to be radial sequence of all the other points p!=p around p. Then, every transition from a point p to q will set the windmill in a particular spot in q. This device tries to clarify that the new point in a windmill sequence depends (only) on the two previous points of the sequence. (Anonymous, 8:41 pm)

3a. If the points form a convex polygon, it is easy. (Anonymous, 8:08 pm)

11a. One can start with any point (since every point of S should be pivot infinitely often), the direction of line that one starts with however matters! (Anonymous, 8:19 pm)

11c. Perhaps even the line does not matter! Is it possible to prove that any point and any line will do? (Anonymous, 8:31 pm)

4a. The first point and line P_0, l_0 cannot be chosen so that P_0 is on the boundary of the convex hull of S and l_0 picks out an adjacent point on the convex hull. Maybe the strategy should be to take out the convex hull of S from consideration; follow it up by induction on removing successive convex hulls. (Haggai Nuchi, 8:08 pm)

23a. Can someone give me *any* other example where the windmill cycles without visiting all the points? The only one I can come up with is: loop over the convex hull of S. (Srivatsan Narayanan, 9:08 pm)

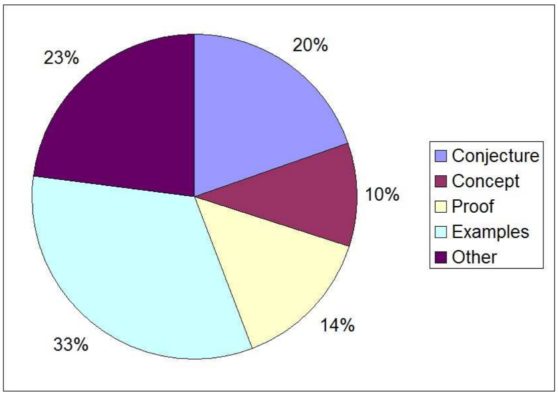

Fig. 1. The proportions of comments which concerned conjectures, concepts, proofs, examples, and other. The comments were those between 8pm and 9.50pm. There was a total of 147 comments, of which 9 were counted twice and 1 was counted three times, since they concerned two/three categories.

We tested the hypothesis that each category would dominate at different stages of the discussion. For instance, it seemed likely that examples would play a prominent role early on, as an aid to understanding the problem, and that concepts would also be discussed early on, and then discussion might turn to proof strategies. Finally, the conjecture might be discussed and its limits explored. However, as can be seen in figure 2, there was little evidence of this: except that comments about concepts did trail off after about the first 45 minutes. In the following four sections, 5 - 8, we consider each category in greater detail, and in section 9 we analyse comments which we categorised as "other".

5 Examples

We see from the pie chart in Figure 1 that examples played a major role. Examples were used for different purposes at different stages of the discussion. One of the first comments (at 8.08pm) was a simple supporting example of the conjecture. This was the only example explicitly raised in this context. Other supporting examples were raised as elaboration (one example) or as highlighting the necessity of a condition in order to explore the condition (two examples). One comment contained an argument as to why a particular example could not exist. There were two requests for clarification, one of an example and the other of a counterexample; and

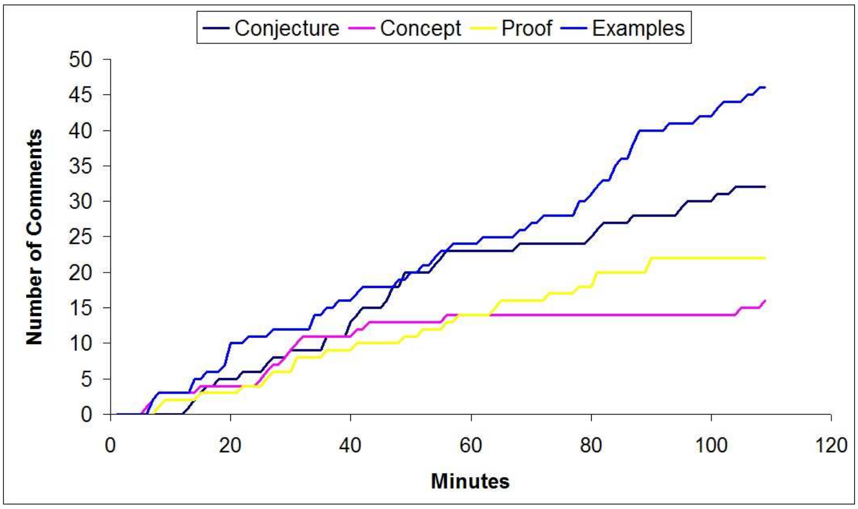

Fig. 2. The time at which the comments concerning conjectures, concepts, proofs and examples were made. The x-axis represents number of minutes past 8pm, July 19th, 2011. The y-axis represents the cumulative number of comments.

three comments which clarified examples, two clarifying examples and the other a counterexample. Of the 48 comments we categorised as concerning examples, 40 were about counterexamples (or examples of undetermined status). Since participants were discussing an IMO problem (which presumably they trusted to be provable), counterexamples were raised tentatively. There were three requests for counterexamples to a sub-conjecture which tested the limits of the given problem. The first request triggered the longest thread, thread 23, which contained eighteen comments including the other two requests (the next longest thread was 14 which produced the proof and had fourteen comments: all of the other threads contained ten or fewer comments with an average of 3.6 comments.[2]) Nine comments discussed “counterexamples” to the initial conjecture: of these three were a class of counterexample and six were specific counterexamples. Fifteen counterexamples were given to sub-conjectures which arose in the discussion: again we can divide these into classes of counterexample and specific ones (seven and eight comments respectively). Eleven comments were responses to counterexamples: all arguing that they are not valid counterexamples, with ten of the eleven giving an explanation as to why not.

5.1 Thread 3

The third thread contains an interesting stream of reasoning. It begins with the observation of a simple class of supporting examples (sets of points which form a convex polygon); this class is then very slightly altered to form the first apparently problematic objects (a set of points which form a convex polygon with a single point not on the convex hull). A third person then gives (the simplest) example of this kind of set (a set of points which form three corners of an equilateral triangle with a fourth point in the centre) and suggests that this appears to be a counterexample.

3a. If the points form a convex polygon, it is easy. (Anonymous, 8:08 pm)

3b. Yes. Can we do it if there is a single point not on the convex hull of the points? (Thomas H, 8:09 pm)

3c. Say there are four points: an equilateral triangle, and then one point in the center of the triangle. No three points are collinear. It seems to me that the windmill can not use the center point more than once! As soon as it hits one of the corner points, it will cycle indefinitely through the corners and never return to the center point. I must be missing something here (Jerzy, 8:17)

This is then quickly followed up by two comments (posted simultaneously) which argue that the object raised is *not* a counterexample.

[2] This calculation ignores the threads which were simply pingbacks.

3d. This isn't true it will alternate between the centre and each vertex of the triangle. (Joe, 8:21 pm)

3e. No, you're not right. Let the corner points be A, B, C, clockwise, M the center. If you start in M, you first hit say A, then C, then M, then B, then A. (Thomas H, 8:21 pm)

The response is then to redefine sub-concepts in the conjecture, thus making the problem object (the vertices of a triangle and an internal point) a supporting object:

3f. Ohhh... I misunderstood the problem. I saw it as a half-line extending out from the last point, in which case you would get stuck on the convex hull. But apparently it means a full line, so that the next point can be "behind" the previous point. Got it. (Jerzy - 8:31 pm)

Thus, the meaning of the concept of rotating line is changed from a half-line to a full line - it extends out in both directions from its pivot.

5.2 Thread 23

Thread 23 concerns an exploration of the problem. Srivatsan raises the question as to whether there are any examples of infinite cycles which do not visit all points in a set, apart from looping over the convex hull. There is then a brief discussion about an example of many equidistant points on a circle and the central point, which concerns whether this example significantly differs from previous examples (figure (*i*) in figure 3 below)). The second example suggested takes the vertices of a regular pentagon and the central point and specifies how the windmill process should start and then loop (figure (*iii*)). Again, it is questioned (by someone other than the person who posed the question) whether this is really an example of the type requested, since the windmill is still looping over the convex hull, albeit in a different way to previously described. The person who suggested this example then varies it slightly in order to address this concern, and suggests that it is now a valid example (in which the points do not form a convex hull)(figure (*iv*)). The next comment presents a new example – an equilateral triangle with a smaller equilateral triangle inside and a point inside the smaller triangle (figure (*v*)). The contributor defines the windmill process and demonstrates why it is an example of the type requested (focusing on the part in which previous examples were considered to have failed - that the visited points form a convex polygon). Srivatsan (who initially made the request) then questions this example, asking for further details as it seems to him to fail on the convex hull criteria too (figure (*vi*)): this appears to be because Srivatsan is naming the vertices in a clockwise direction (in which case the windmill process would end up on the convex hull, as he points out), as opposed to that assumed by Varun and Seungly Oh, who name them in an anticlockwise direction. Another contributor elaborates on the example to show that it *is* valid, Srivatsan then agrees that it is of the type he requested, and this ends the search.

The next comment in the thread, by Joel, is another request for an example: this is related to the first request, with further criteria which the accepted example, directly above, would fail. He asks whether there is an example where the space not swept by the windmill is unbounded (in the previous example the space is bounded by the internal triangle, and in all examples raised up to now the space not swept is bounded at the most, by the convex hull of the points). The next comment argues that such an example could not exist (any unbounded space outside the convex hull will eventually be swept by the windmill). The following comment seems to be a clarification of the type of example Joel has requested. The response is a clear belief in the conjecture, saying that this is "graphically obvious" and asking how this helps (presumably, helps to solve the initial conjecture). The next response is agreement with both points. A final example of the first question in thread 23 is introduced, via the initial example raised in this thread. This example is slightly modified by perturbing some of the points (figure (*ii*). Srivatsan, who began the thread, thanks contributors for their examples.

6 Conjectures

Twenty nine comments concerned conjectures. This included the initial conjecture and exploration of its limits, and sub-conjectures which were made, often in conjunction with a proof attempt (we could equally have categorised these comments as concerning proof). There was one correction to the problem statement (this was acknowledged and addressed by Tao who set the problem); one clarification of the problem and three clarifications of a property or sub-conjecture; one comment rephrased the initial conjecture; two generalised the

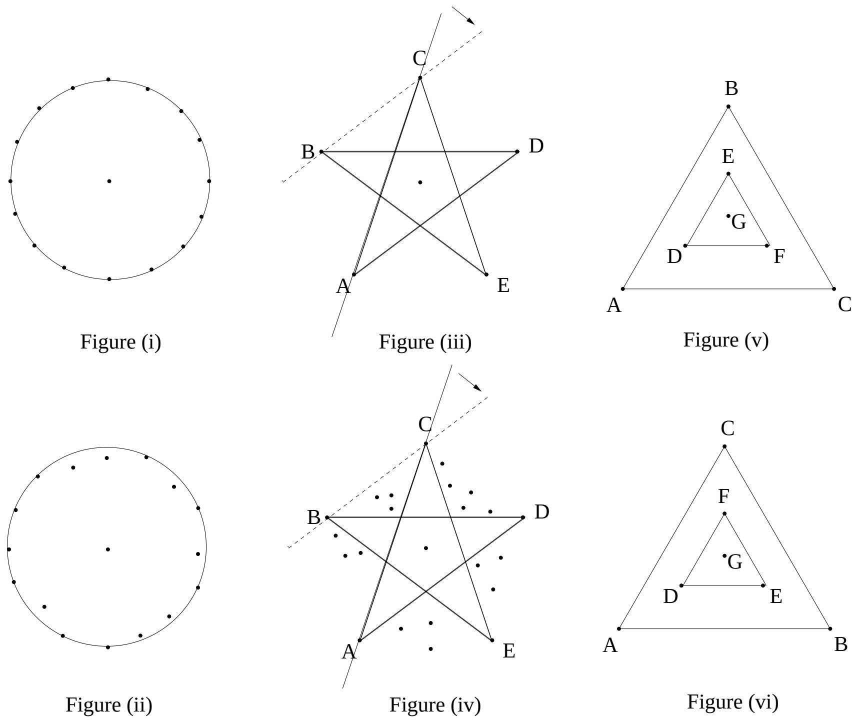

Fig. 3. Examples raised in thread 23. Figure (*ii*) is a variation on figure (*i*); figure (*iv*) is a variation of figure (*iii*); and figures (*v*) and (*vi*) are the same figure, where one is labelled clockwise and the other anti-clockwise.

initial conjecture. Eighteen comments asked, or proposed sub-conjectures (normally related to proof); and three responded to sub-conjectures or a generalised form of the initial conjecture.

Conjectures were suggested in a variety of ways, ranging from tentatively asking it (for example 11c, 19a below), to proposing (2g below) or even stating it (13c below). This perhaps reflected the confidence level that the proposer had in their conjecture. We categorise four types of conjecture, giving examples of each below. Note that most comments on thread 14 can be found in section 8.3.

2g Is it me, or you can get a partition of the set of directed edges on this graph into admissible cycles (i.e., cycles generated by the windmill process) ? You juste have to reverse the time if necessary (Garf, 8:37 pm)

11c Perhaps even the line does not matter! Is it possible to prove that any point and any line will do? (Anonymous, 8:31 pm)

13c There must be n/2 distinct cycles since the number of points to the left or right of the line remains constant throughout the process. (Correct me if I'm wrong.) (Justin W Smith, 8:50 pm)

19a A question: Does the windmill process eventually form a cycle? (Seungly Oh, 8:48 pm)

---

1. Conjectures made by analogy to another domain. After an analogy to another domain has been proposed (usually by someone else) and concepts defined (eg 2g), the initial conjecture is translated, or new sub-conjectures are formed, in the new domain. This is done in graph theory (2g above), also in terms of angles (thread 9, 22) and functions (thread 17).
2. Conjectures which extended the initial problem. The initial conjecture was changed slightly when participants realised that the starting point does not matter (in thread 11 and also comment 14c). This generalises the conjecture, since the problem only requires one to prove that there *exists* a point P in S, instead the participants prove the much stronger theorem *for all* points P in S. The question is then also raised as to whether the choice of line matters (11c above), and the number of distinct cycles is conjectured to be n/2 in thread 13.

3. Sub-conjectures towards a proof. Sub-conjectures, or observations, which may be relevant to a proof are formed, such as the conjecture that it is possible to draw a line through any point which divides the remaining points into to roughly equal sets (thread 13c above and 14a). A second example is the observation that the number of points to left or right stays constant throughout the windmill process (14d).
4. Conjectures about the main concept. There were five conjectured properties of the windmill process, in threads 18, 19a above and 23. We consider these further in section 7.2.

7 Concepts

Concepts are introduced and developed in a variety of ways. These include drawing analogies to other domains, correcting misunderstandings, using standard concepts which would be known to people familiar with the field, using the conjecture to generate concepts, using a proof idea to generate concepts and introducing conventions which allow new concepts to be defined in a particular way. We discuss some of these below (7.1 - 7.4), and in section 8.2 we discuss proof-generated concepts.

7.1 Analogies

Both conceptual metaphors to everyday objects and analogies to structurally similar domains are made and used during the discussion. The key concept in the conjecture, *windmill*, is itself a metaphor, and Geoff Smith, the problem's author, explains that the fact that the 2011 IMO was held in the Netherlands led him to think of Dutch icons, which were the inspiration behind his invention of the windmill concept:

> "The problem is based on the idea of a rotating line, and was specifically designed as a windmill problem for the Dutch IMO." [39, p. 2]

Another metaphor to an everyday object is invoked in comment 15a as a proof strategy:

> "We could perhaps consider "layers" of convex hulls (polygons) .. like peeling off an onion. If our line doesn't start at the "core" (innermost) polygon then I feel it'll get stuck in the upper layers and never reach the core." (Varun, 8:27pm)

The use of analogy in Mini-Polymath is a tool for driving discovery as well as playing an explanatory role. A search on the term "like a" produces three results:

2a. Connecting the dots: At the point where the pivot changes we create a line that passes through the previous pivot and a new pivot like a side of a polygon. (Gal, 8:07 pm)

2f. Or like an edge of a graph, and each edge leads to another edge. We want to show that there's a circuit that visits every vertex at least once. Ideas? (Anonymous, 8:28 pm)

14k. Got it! Kind of like a turn number in topology. Thanks! :) (Gal, 9:50 pm)

The first comment above, 2a, introduces terminology which is then widely used throughout the discussion: seeing (part of) the set of points as a polygon. The word "polygon" appears 15 further times in the discussion, and is important for expressing examples and counterexamples (the concept "convex polygon" is particularly important) as well as for suggested proof strategies. While we could debate whether this is really an analogy, it certainly furnishes the group with conceptual apparatus for discussion the set of points S in the question. The second comment, 2f, translates the conjecture and relevant concepts into the domain of graph theory, which is then taken up by other participants in various proof attempts. In thread 13 the same analogy is made and explored via examples and observations (seemingly independently, with the first comment at 8:26 pm). One comment explicitly addresses the problem of mapping between the terminology of the question and graph theory:

> The question here is how to translate the inherent geometrical properties that are required to prove the statement into properties of the graph. (Since obviously it isn't true for all graphs) (Joe, 8:30 pm)

The third comment above, 14k, is explanatory: the participant is showing his understanding of a concept (which simultaneously explains the concept to others).

One analogy which is not taken up collaboratively is in thread 12, which contains three comparatively detailed and lengthy comments, all by the same author, exploring an analogy to projective duality. Despite the first comment being positively rated twice, it is not explored by other participants: this may be due to the relative obscurity of the target domain.

7.2 Conjecture-generated concepts

Concepts are developed in order to describe properties of key concept in the conjecture: the windmill process. For instance, one property (which turned out to be important) was discovered early on: it is time reversible, that is, there are no processes which have an initial segment which does not repeat (2h). Another property relates to angles, and provides a way of finding the next pivot by measuring angles (9a).

7.3 (Counter)example-generated concepts

Concepts develop in order to describe a class of examples or counterexamples, as properties of an example, and as responses to counterexamples or misunderstandings.

Concepts arise in the context of defining a class of object (which might be examples or counterexamples to a conjecture). These gradually became more complex, and often appeared to be constructed from simpler classes which has arisen in the discussion. For instance, in comment 3a the class of example: *a set of points which form a convex hull* is introduced. In 3b we then see a slight variation on this: *a set of points which form a convex hull with a single interior point*. Likewise, in 23b we see the class *many equidistant points which lie on the circumference a circle, and a central point* defined, and later, in 23q, the variation *many roughly equidistant points which lie roughly on the circumference a circle, and a central point*. Sometimes a procedure for constructing such a class was also supplied (as in the case of 23q).

Another conceptual extension is giving a direction to the line, which allows us to deduce useful properties: a direction means that it makes sense to talk of points being to the left or the right of the line. Participants then realised that the number of points to the left or right of the line stays constant throughout the windmill process.

We also see conceptual development as a result of responding to counterexamples. In section 5 we saw one such development, in which the concept *line* was clarified as full rather than half.

7.4 New conventions

New concepts are defined by setting and following conventions. There were 31 instances of participants using quotation marks, usually signalling the use of a specific word or concept in a new context. Once defined in the context, other participants often followed the convention, using the word in the same way. For instance, in thread 23 we see the word "sweep" being used in a very particular context: this word is then used in the following two comments as well as later on in the discussion (the discussion is described in section 5.2). Some of these contain key insights, such as a reference to "splitting" lines in thread 13, which was fundamental to the proof. Others refer to properties not usually associated with a concept, such as comment 14d which talks about the "left" or "right" of a line (this is another key insight) and 14e which talked about a line being "upside-down". Similarly, in thread 23 one participant refers to "harmless" points: harmfulness is not usually a property of points, but in the context the meaning is clear. Quotation marks are also used to expand metaphors: in comment 15a (shown above in section 7.1) we see quotation marks used to expand the metaphor to an onion.

8 Proofs

Twenty one comments concerned a proof. Fourteen were about proof strategy, one was clarification of the strategy, one was carrying out a plan and five were about identifying which properties were relevant to the proof.

8.1 Proof strategies

Various proof strategies are found. Key ideas include (*i*) induction on the convex hulls which the points form (comments 2 and 4); (*ii*) using an analogy between Euclidean space and projective geometry (comment 12), where a correspondence between a *point p* (in Euclidean space) is drawn between *a line m_p in an open Möbius strip M* (in the projective plane) and likewise a *line l passing through a point p* is mapped to a *point n_l on m_p in M*. Appropriate correspondences are then made for the windmill process and the goal; (*iii*) another analogy is proposed, to directed graphs (comment 13), by slightly re-representing the set S as a set of ordered pairs of points, where each point represents a transition from one pivot point to the next. Again, key concepts such as members of this new set S and the windmill process are mapped to appropriate concepts in the target domain and the problem is translated.

8.2 Identifying relevant properties

Exploring the properties of the windmill process and identifying which of them are relevant, lead to the introduction of proof-generated concepts. For instance, in comments 11g, 11i and 11j, participants discuss how to select a starting point and line, the relationships between them, and which choices matter in terms of finding a windmill process. Similarly, in 17g conditions of the conjecture are weakened, and in 18a a new concept, the "wheel of p" is introduced in order to demonstrate the relationship between the next point in the windmill sequence and the previous two points.

8.3 Analysis of thread 14

In the 14th thread we see the emergence of the solution which is recognised by Tao. It is a fine example of collaboration, with between seven and nine people contributing to the 16-comment thread (one of whom is "Anonymous", who comments three times). Key insights are provided by three participants (Anonymous, Justin W Smith and Garf), four participants (Thomas H, Gal, Zhecka and Jerzy) encourage that the line of reasoning looks fruitful, and two participants (Gal and Jerzy) analyse and further clarify the proof. Below we consider the thread up to 14k: there are five further comments in the thread (two of which come *after* Tao has commented that he considers it to be a valid solution, at 9.50pm).

14a. I'm not sure but as no three points are collinear, one can always find a line which splits the points into two sets whose number of elements differ at most one? (Anonymous, 8:27 pm)

14b. That is surely true. How could this help us? (Thomas H, 8:28 pm)

14c. Something like one can find this no matter how we choose the first point. Then in some time the windmill must be parallel to the line through these points. This line must be unique or else it splits the points such that number of elements differ at least two. (Anonymous, 8:41 pm)

14d. It appears that the number of points to the "left" or "right" of the line is constant through the entire process! (Justin W Smith, 8:47 pm)

14e. I think this solves the problem. Start with a line which separates the points into two parts of roughly same size (their cardinal differ by at most one, not counting the point to which the line is attached). Then run the process until the line is "upside-down", and so has turn by exactly π. Every point has gone from the right of the line to the left of the line (easy to see is the number of point is odd, you have to be a bit more crafty if it is even), and no point can go from left to right or right to left without touching the line. Add the previous remarks (the process will always come back to its initial configuration), and every point will be visited infinitely often. (Garf, 9:14 pm)

The first comment (14a) is an observed conjecture about a property which holds about a concept in the conjecture. It is not (necessarily) immediately apparent as to how this observation will be relevant to a proof. The second (14b) recognises that the conjecture holds and asks how it might be relevant, implicitly adding value to the first comment by the response. In the third (14c) participants generalise the problem, proving the much stronger theorem *for all* points P in S, rather than the weaker stated problem which only requires one to prove that there *exists* a point P in S. A second conjecture is introduced "Then in some time the windmill must be parallel to the line through these points" and then a case split argument is used on the line. 14d is similar to 14a - introducing a new conjecture about properties of a key concept in the problem (in this case, about the windmill; in 14a the concept was the set S). 14e contains a procedural proof of the problem. It is interesting to note the ambiguity of the word "this" in the first sentence: it may refer to what has just gone before, or the argument which it precedes.

14f. Very nice! Don't we run into problems with a convex hull though? Take a square with a point in the middle (M) and pass the diagonal of the square (not through M) it seems to me M is never visited (though I may be wrong here). I think we should be more specific in our initial choice of line, maybe? (Gal, 9:23 pm)

14g. No. This example is false :) (Gal, 9:28 pm)

14h. Yes, it seems to be a correct solution! (Zhecka, 9:35 pm)

14i. This seems to be right, but there something I don't understand. Please see if you can help me with it:
Start with a square and a point inside it (M): start with a tangent to the square (your solution demands a more equal division of points, I know). When we get to the opposite vertex of the square all points moved from one side of the line to the other, but not all points have been visited (M will never be visited). The argument is almost exactly the same, so it seems that the equal division of points plays a crucial role, but I dont understand what role exactly. Can we pin it down precisely? (Gal, 9:42 pm)

14j. If I understand well your example : the problem is that you must give an orientation to the line. Then, left and right are define with respect to this orientation : if the line has made half a turn, then left and right are reversed. In your example, I think most of the point move from, say, the part et the top of the line to the part at the bottom of the line, but always stay at the right of the line. (Garf, 9:47 pm)

14k. Got it! Kind of like a turn number in topology. Thanks! :) (Gal, 9:50 pm)

14f is a suggested (specific) counterexample to the proof, which five minutes later the author of the comment withdraws, with the explanation that it is "false". Certainly, it looks as though this was purely a mistake, since M *will* be visited. It is clear that the author continues to think about his example, however, since - after recognition that the proof contained in comment 14e seems to be correct (14h), the participant raises it again in 14i. This is in the context of understanding the role which the condition of dividing the set into two roughly equal sets of points plays: it is acknowledged to be important since the (slightly changed) "counterexample" which does not satisfy the condition appears to fail the proof (since not all points will have been visited). This plea for explanation is answered in a way which is (presumably) different to that expected (14j): it results in a clarification of the concepts "right" and "left of the line", rather than an examination of the condition. It clarifies the problem to the author however, who draws an analogy to topology to show, or extend his understanding of the concept (14k).

9 Other

The 34 comments we classified as "other" can be broken down into roughly three subcategories. There was some overlap between comments, with duplication of examples, proof strategies, and so on. One important contribution which people made was to cross reference when this happened, back to the initial thread: there were 11 such comments. In seven comments participants explained a claim. There were 14 comments which we classified as meta-comments (discussed below). The other two comments concerned expansion and justification (1 comment in each category).

Contributors made evaluative comments on aspects raised in other comments. These took the form of showing appreciation, either generally for a good idea, or for a response to a specific question (for example, 2b and 23r below); applauding areas which look fruitful (15b); confirming that someone has understood correctly (14n), or confirmation of a correct solution (14h) (there was also a ratings system whereby each comment could be positively or negatively rated). There were also comments on a proof strategy, such as reasons why a suggested proof strategy might be difficult (15c), or simply asking how an observation might help with a proof (14b). These follow normal rules of discourse, in which friendly interjections are made: some of which are mathematically interesting, guiding the direction of the discussion, while others are simply courtesy comments (below, we would classify 2b, 15b, 14n, 15c, 27a, 14b as the former and 23r the latter). All of them play an important role in keeping the conversation flowing. Participants also use of smiley faces, exclamation marks, and so on, which all contribute to creating and maintaining an environment which is friendly, collaborative, informal, polite and psychologically safe.

2b. Nice... Garf, 8:23 pm

23r. AT Thomas AT Seungly AT Haggai Thank you all for your examples. I haven't understood them fully yet; I'll think about them for some time and get back if I have questions. (Srivatsan Narayanan 9:39 pm)

15b. I think that is a good start, thanks Varun! (A, 8:46 pm)

14n. That's it. (Garf, 9:50 pm)

14h. Yes, it seems to be a correct solution! (Zhecka, 9:35 pm)

15c. This was suggested by Haggai Nuchi (comment 4). But its hard to keep track of the process since it keeps switching between multiple layers in a seemingly arbitrary fashion (in fact, we are looking for a windmill that does exactly this). (Srivatsan Narayanan, 8:48 pm)

14b. That is surely true. How could this help us? (Thomas H, 8:28 pm)

10 Theories of mathematical practice

Pólya and Lakatos were the forebears of a body of research which lies at the junction of the history, philosophy and cognitive science of mathematics and includes work on visualisation, metaphors, analogies, concept blends, network theory, concepts and definitions, heuristics for discovery and justification, and social aspects of mathematics (see, for instance, [9, 14, 27]).

10.1 Pólya's problem-solving heuristics

Pólya characterised problem-solving methods, collected in his *How to Solve it* [36], in order to aid the teaching and learning of mathematics. The career of these heuristics has been somewhat chequered, peaking early on with a profusion of problem-solving courses and seminars in which [36] featured as a textbook, and then declining somewhat as it became apparent to would-be mathematicians and their educators that while possibly being necessary, these heuristics were certainly not sufficient and did not replace experience (see Ian Stewart's foreword in the 1990 edition of [36]). One key insight regarded the lack of meta-heuristics for determining when each heuristic might be fruitful or when it would be more productive to abandon one line of thought and search for an alternative strategy [38]. Whatever the view of their utility, given the educational setting, we must be wary of circularity when analysing the discussion from a heuristic point of view. Pólya's problem solving heuristics do seem to shine through: for example we see participants rephrasing the question (comment 2), using case splits (comments 5 and 14) and trying to generalise the problem (comments 14 and 23). This is hardly surprising, as the questions themselves may have been written and judged depending on whether they can be solved with this way of thinking. The Mini-Polymath context does, however, present an interesting opportunity to see how Pólya's heuristics operate in a collaborative, as opposed to individual, setting.

10.2 Lakatos's patterns of reasoning

Lakatos presented a fallibilist approach to mathematics, in which proofs, conjectures and concepts are fluid and open to negotiation [24]. He criticised the deductivist approach in mathematics, which presents definitions, axioms and theorem statements as immutable ideas which come from nowhere into a mathematician's empty mind. Instead, he outlined a heuristic approach which holds that mathematics progresses by a series of primitive conjectures, proofs, counterexamples, proof-generated concepts, modified conjectures and modified proofs. The Polymath and Mini-Polymath projects afford precisely the sort of openness that he advocated in the teaching and presentation of mathematics. Lakatos categorised responses to, and uses of, supporting and counterexamples to describe the conversation. He emphasised fallibility and ambiguity in mathematical development, addressing semantic change in mathematics as the subject develops, the role that counterexamples play in concept, conjecture and proof development, and the social component of mathematics via a dialectic of ideas. Although his theory was highly social, it was not necessarily collaborative. He was describing research mathematics in which examples, conjectures, concepts and proofs were all on the examination table and open to discussion: in Mini-Polymath the conjecture at least enjoys a special status as something unchangeable. Thus, Lakatos's methods which result in rejection or refinement to a conjecture under discussion, namely surrender, his exception-barring methods and global lemma-incorporation, might seem to be irrelevant (although we can see them applied to sub-conjectures). We do see examples of those methods which result in conceptual change – monster-barring and monster-adjusting – and local lemma-incorporation, which results in refinement to a proof plan.

Monster-adjusting occurs when an object is seen as a supporting example of a conjecture by one person and as a counterexample by another; thus exposing two rival interpretations of a concept definition. The object then becomes a trigger for concept development and clarification. We see an example of this in comment 3 (Section 5.1) in which the problematic object is an equilateral triangle with one point in the centre; this exposes different interpretations of the concept of rotating line, as either a half-line (extending from a pivot) or a full-line. The issue is resolved almost immediately. While Lakatos identifies the role that hidden assumptions play, and suggests ways of diagnosing and repairing flawed assumptions, he does not suggest how they might arise. Here we can go beyond Lakatos and hypothesise as to what might be the underlying reason for mistaken assumptions or rival interpretations. Lakoff and colleagues [25] and Barton [7] have explored the close connection between language and thought, and shown that images and metaphors used in ordinary language shape mathematical (and all other types of) thinking. We hypothesize that the misconception of a line as a half-line may be due to the naming of the concept; which triggered images of windmills with sails which pivoted around a central tower and extended in one direction only.[3]

The main case study in [24] starts with a proof which can be seen as a set of procedural steps P_i combined with a set of declarative facts D_i, of the form: Do P_1, then D_1 holds; Do P_2, then D_2 holds; Do P_3, then D_3 holds. Following the chain of reasoning back up through the proof then gives $D_3 \rightarrow D_2 \rightarrow D_1 \rightarrow$ Theorem. Problems or counterexamples are either of the form "How do you know you can always do P_i?" or "I have done P_i but D_i doesn't hold". The revision for the latter is to use the counterexample to suggest further conditions and then replace P_i by P_i', where carrying out step P_i' on the counterexample *does* result in D_i being the case. In comment 14 (see 8.3) we see a similar structure in the argument, consisting of a set of procedures to follow (P_i) and declarative facts about the procedures, although with some important differences. The simple, and what turns out to be crucial, observation in 14a that there is a line through any point which divides the points into two sets whose number of elements differ by one at most, corresponds to one can always do P_i, which results in a participant acknowledging that that is the case but asking how is it relevant. The resulting comments can be expressed in terms of the Lakatosian structure above in the following way: One can always do P_1 (find a line which splits the points into two roughly equal sets). Do P_2 (the windmill procedure). Then, at some point, D_{2i} and D_{2ii} will hold (the line will be parallel to the initial line and every point has gone from the right of the line to the left of the line). This chain implies the Theorem. In Lakatosian fashion, this proof then opens up the target for counterexamples, which are duely proposed.

The two case studies which Lakatos considers concern mathematics at a very high level, based on work by Cauchy, Gergonne, Poincaré and other mathematical giants. Therefore, it may be unfair to expect the same logic to apply to (presumably) everyday mathematicians or mathematics students: in particular in the mini-polymath examples in which discoveries made will be psychologically-creative (new to an individual) rather than historically-creative (new to a domain). In his second appendix, *The deductivist versus the heuristic approach* [pp. 142 - 154], Lakatos argues, via an impassioned critique of what he terms "Euclidean methodology", that students of mathematics would greatly benefit from a more honest presentation of material. Rather than making her way through a baffling series of unexplained definitions, theorems and proofs, the student would learn more if the true struggle and adventure concealed by such a bland presentation were revealed. This attitude somewhat conflicts with Lakatos's intention of proving helpful heuristics for mathematical discovery, and his selection of case study. Except for surrender, all of his methods – even the so-called "heuristically sterile" ones which are seen as primitive and unhelpful – result in some small step of progress. Even if there are more constructive ways of reasoning, something further is known after the application of the method than before. In reality, much mathematical thinking is simply flawed. Objects thought to be counterexamples are barred – not because of any interesting definitional change in key concepts – but because of boring mistakes; mistakes in calculations or misunderstandings which may have led to an individual's enhanced understanding, but certainly did not add anything to the field (see, for instance, comments 14f and 14g above). Here we see that oft-discussed blurring of motivation in Lakatos's work between prescriptive and descriptive. Since he formed his theory from hand-picking case studies and the major steps of development made within each case study, we could accuse Lakatos himself of hiding the struggle and the adventure in his own formation of a logic of discovery. An indepth study of the Polymath projects should enable us to form theories of mathematical reasoning which are far more descriptive of everyday thinking than Lakatos's theory.

[3] The IMO presents tremendous opportunity for cultural and linguistic analysis, as each problem is translated into at least five different languages, and candidate problems are evaluated partially for the ease with which they can be translated, and the process of translating a problem is taken extremely seriously.

11 A computational blogger

Turing formalised the notion of algorithm [41], and algorithms for finding proofs have been defined and implemented as computer programs in a host of theorem provers, such as Otter [31], Vampire [37] and SPASS [8]. Gowers [20] asks whether the notion of a "good proof" could similarly be formalized, arguing that such a formalization would have as great an impact on mathematics as the formalization of algorithm and proof. He considers some characteristics of good proofs, namely understandability and explanatory power, and proposes (in true Lakatosian fashion) a rational reconstruction of mathematical results, in order to demonstrate the origin of concept definitions, proofs and conjectures. In particular, Gowers suggests looking at working methods of mathematicians, as a useful first step to teaching computers (and students) how to do mathematics:

> ... "if we wish to teach computers to find proofs, it is likely to be a good idea to reflect on how we do so ourselves." [20, p. 4]

In a seemingly independent project, his Polymath experiments, Gowers has provided the very means for doing just that: the working methods and informal thought processes of mathematicians are recorded, forming a body of data which is posted online in a searchable record. This clearly holds much potential for illuminating mechanisms behind human mathematical thought.

We have analysed the Mini-Polymath discussion in terms of whether the focus of a comment was an example, concept, conjecture, proof or something else. We can see automated systems in the same light. Systems under each category include:

- Examples: Model generators find examples; for instance MACE [29] searches for finite models of first-order statements;
- Concepts and conjectures: Automated theory formation systems automatically invent concepts and conjectures. Examples include Lenat's AM [26], which was designed to both construct new concepts and conjecture relationships between them; Colton's HR system [11, 12], which uses production rules to form new concepts from old ones, employs a set of measures of interestingness to drive a heuristic search, uses empirical pattern-based conjecture making techniques to find relationships between concepts, and employs third party automated reasoning systems to prove the conjectures or find counterexamples. Other examples include McCasland's MATHsAiD project [28], which aims to build a tool which takes in a set of axioms, concept definitions and a logic and applies its inference rules to reason from the axioms to theorems; the IsaScheme system by MontanoRivas [32], which employs a scheme-based approach to mathematical theory exploration, and the IsaCosy system by Johansson *et al.* [22] which performs inductive theory formation by synthesising conjectures from the available constants and free variables;
- Proofs: Theorem proving systems and computational logic systems are the best known of the automated mathematicians. The most famous examples of computers being used in a proof are Appel and Haken's claim that they had proved the Four Colour Theorem with extensive help from a computer (see [42] for a history of the proof and controversy), and computational proofs of the Robbins Conjecture by McCune [30] and the Kepler Conjecture [21] by Hales and Ferguson.

A handful of systems are built on theories of mathematical practice. Some examples of systems which are based on work by Pólya and Lakatos include the following:

- Pólya: Pólya's work on problem solving [36] has inspired computational accounts, such as emulations of his work on reasoning by analogy [16, 23], a Pólya-inspired explanation of algorithms and data structures [15] and an account of how to select between different problem-solving methods [17];
- Lakatos: The HRL system by one of us [34] is a computational account of Lakatos's theory [24], implemented in an agent architecture where each agent has a copy of the HR system [11]. Agents form and communicate conjectures, and then find and respond to counterexamples, following Lakatos's theory. Colton and Pease's TM system [13] is also based on some of Lakatos's methods. This takes a set of axioms and a conjecture in first order logic, attempts to prove it, and if it fails, attempts to modify the conjecture into a theorem which it can prove.

It is clear that, despite many successes in automating various aspects of mathematical theories, we are still a very long way from a system which could contribute, in a human-like manner, to a Mini-Polymath discussion. What more would it take? Progress could be made in two directions. Firstly, we could combine systems which form analogies, generate natural language, and so on, with the automated model generation, theory formation and theorem-proving systems we have described. Secondly, we could further analyse mathematical discussions in order to determine when, why and how particular aspects are raised, and use this to develop systems which specialise in one aspect. In either case further analysis of the discussion in the Polymath and Mini-Polymath projects would be enormously fruitful. Our analysis of the third Mini-Polymath discussion has shown that participants raise and discuss examples and counterexamples, propose conjectures and express a level of confidence in them, extend an initial problem; form analogies to other domains and translate conjectures and concepts from one domain to another; explore properties of the main concept in the initial conjecture; correct misunderstandings; find and clarify proof strategies; evaluate other people's contributions; cross-reference previous comments; and perform a multitude of other tasks, all the while following appropriate conventions of discourse, such as use of other participants' names, friendly interjections, smiley faces, exclamation marks, etc. Undoubtedly, Turing [40] was right when he commented: "We can only see a short distance ahead, but we can see plenty there that needs to be done.".

Bibliography

[1] General polymath rules. http://polymathprojects.org/general-polymath-rules/.

[2] Mini-polymath3 discussion. http://terrytao.wordpress.com/2011/07/19/mini-polymath3-discussion-thread/.

[3] Minipolymath3 project. http://polymathprojects.org/2011/07/19/minipolymath3-project-2011-imo/.

[4] The polymath blog. http://polymathprojects.org/.

[5] IMO. http://michaelnielsen.org/polymath1/index.php?title=Imo_2011, 2011.

[6] B. Baars. *A cognitive theory of consciousness*. Cambridge University Press, 1988.

[7] B. Barton. *The Language of Mathematics: Telling Mathematical Tales*. Mathematics Education Library, Vol. 46. Springer, 2009.

[8] Weidenbach C., Dimova D., Fietzke A., Kumar R., Suda M., and Wischnewski P. SPASS version 3.5. *22nd International Conference on Automated Deduction*, pages 140–145, 2009.

[9] C. Cellucci and D. Gillies. *Mathematical reasoning and heuristics*. Kings College Publications, London, 2005.

[10] J. Charnley and S. Colton. A global workspace framework for combining reasoning systems. In *Proceedings of the Symposium on the Integration of Symbolic Computation and Mechanised Reasoning (Calculemus)*, 2008.

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

[12] S. Colton and S. Muggleton. Mathematical applications of Inductive Logic Programming. *Machine Learning*, 64:25–64, 2006.

[13] S. Colton and A. Pease. The TM system for repairing non-theorems. In *Proceedings of the International Joint Conference on Automated Reasoning (IJCAR) workshop on Disproving*, 2004.

[14] D. Corfield. Assaying Lakatos's philosophy of mathematics. *Studies in History and Philosophy of Science*, 28(1):99–121, 1997.

[15] R. G. Dromey. *How to Solve It by Computer*. Prentice Hall, 1982.

[16] T. G. Evans. A program for the solution of geomtric-analogy intelligence test questions. In M. Minsky, editor, *Semantic Information Processing*, pages 271–353. The MIT Press, Cambridge, Massachusetts, 1968.

[17] E. Fink. How to solve it automatically: Selection among problem-solving methods. In *Proceedings of Artificial Intelligence Planning Systems*. AAAI, 1998.

[18] T. Gowers. Polymath1 and open collaborative mathematics. http://gowers.wordpress.com/2009/03/10/polymath1-and-open-collaborative-mathematics/.

[19] T. Gowers. Is massively collaborative mathematics possible? http://gowers.wordpress.com/2009/01/27/is-massively-collaborative-mathematics-possible/, March 2009.

[20] W. T. Gowers. Rough structure and classification. *GAFA (Geometric And Functional Analysis)*, Special volume – GAFA2000(1–0), 2000.

[21] T. C. Hales and S. P. Ferguson. A formulation of the kepler conjecture. *Discrete and Computational Geometry: an International Journal of Mathematics and Computer Science*, 36(1):21–69, 2006.

[22] M. Johansson, L. Dixon, and A. Bundy. Conjecture synthesis for inductive theories. *Journal of Automated Reasoning*, Forthcoming, 2010.

[23] R. E. Kling. Reasoning by analogy with applications to heuristic problem solving: A case study. Technical report, Stanford Artificial Intelligence Project Memo AIM-147, CS Department report CS-216, August, 1971.

[24] I. Lakatos. *Proofs and Refutations*. CUP, Cambridge, UK, 1976.

[25] G. Lakoff and R. Núñez. *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*. Basic Books, 2000.

[26] D. B. Lenat. Automated theory formation in mathematics. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, pages 833–842, Cambridge, MA, 1977. Morgan Kaufmann.

[27] P. Mancosu (Ed). *The Philosophy of Mathematical Practice*. Oxford University Press, 2008, USA.

[28] R. McCasland and A. Bundy. MATHsAiD: a mathematical theorem discovery tool. In *Proceedings of SYNASC*, 2006.

[29] W. McCune. A Davis-Putnam program and its application to finite first-order model search. Technical Report ANL/MCS-TM-194, Argonne National Laboratories, 1994.

[30] W. McCune. Solution of the Robbins problem. *Journal of Automated Reasoning*, 19(3):263–276, 1997.

[31] W.W. McCune. Otter 3.0 Reference Manual and Guide. Technical Report ANL-94/6, Argonne National Laboratory, Argonne, USA, 1994.

[32] Omar Montano-Rivas, Roy McCasland, Lucas Dixon, and Alan Bundy. Scheme-based theorem discovery and concept invention. *Expert Systems with Applications*, (0):–, 2011.

[33] M. Nielsen. *Reinventing Discovery: The New Era of Networked Science*. Princeton University Press, USA, 2011.

[34] A. Pease. *A Computational Model of Lakatos-style Reasoning*. PhD thesis, School of Informatics, University of Edinburgh, 2007. Online at http://hdl.handle.net/1842/2113.

[35] A. Pease and Colton S. On impact and evaluation in computational creativity: A discussion of the Turing test and an alternative proposal. In *Proceedings of the AISB symposium on AI and Philosophy*, 2011.

[36] G. Pólya. *How to solve it*. Princeton University Press, 1945.

[37] A. Riazanov and A. Voronkov. The design and implementation of VAMPIRE. *AI Comm*, 15(2-3):91–110, 2002.

[38] A.H. Schoenfeld. Pólya, problem solving, and education. *Mathematics Magazine*, 60(5):pp. 283–291, 1987 (Dec.).

[39] G. Smith. Observations: The 2011 IMO in the Netherlands. Technical report, 2011.

[40] A. M. Turing. Computing machinery and intelligence. *Mind*, 59:433–460, 1950.

[41] A.M. Turing. On computable numbers, with an application to the entscheidungs problem. *Proceedings of the London Mathematical Society*, 42(2):23065, 1937.

[42] R. Wilson. *Four Colors Suffice*. Penguin Books, London, 2002.