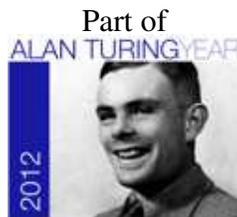


# AISB/IACAP World Congress 2012

Birmingham, UK, 2-6 July 2012

## Symposium on Mathematical Practice and Cognition II

A. Pease and B. Larvor (Editors)



Published by  
The Society for the Study of  
Artificial Intelligence and  
Simulation of Behaviour

<http://www.aisb.org.uk>

ISBN 978-1-908187-10-9

## Foreword from the Congress Chairs

For the Turing year 2012, AISB (The Society for the Study of Artificial Intelligence and Simulation of Behaviour) and IACAP (The International Association for Computing and Philosophy) merged their annual symposia/conferences to form the AISB/IACAP World Congress. The congress took place 2–6 July 2012 at the University of Birmingham, UK.

The Congress was inspired by a desire to honour Alan Turing, and by the broad and deep significance of Turing's work to AI, the philosophical ramifications of computing, and philosophy and computing more generally. The Congress was one of the events forming the Alan Turing Year.

The Congress consisted mainly of a number of collocated Symposia on specific research areas, together with six invited Plenary Talks. All papers other than the Plenaries were given within Symposia. This format is perfect for encouraging new dialogue and collaboration both within and between research areas.

This volume forms the proceedings of one of the component symposia. We are most grateful to the organizers of the Symposium for their hard work in creating it, attracting papers, doing the necessary reviewing, defining an exciting programme for the symposium, and compiling this volume. We also thank them for their flexibility and patience concerning the complex matter of fitting all the symposia and other events into the Congress week.

John Barnden (Computer Science, University of Birmingham)  
Programme Co-Chair and AISB Vice-Chair  
Anthony Beavers (University of Evansville, Indiana, USA)  
Programme Co-Chair and IACAP President  
Manfred Kerber (Computer Science, University of Birmingham)  
Local Arrangements Chair

## Foreword from the Symposium Chairs

At the AISB convention 2010 in Leicester, the multidisciplinary symposium Mathematical Practice and Cognition welcomed researchers into mathematical practice from cognitive science, philosophy, psychology, computational linguistics and robotics. In this second symposium of the series, we continue the trend. This volume includes work in the philosophy, history and sociology of mathematics, as well as argumentation theory, artificial intelligence and computer science.

In the first paper, Andrew Aberdein argues that proof consists of the twin elements of informal argumentation, whose goal is to persuade fellow mathematicians of a result, and rigorous proof, consisting of valid mathematical inferences. He then uses this distinction to suggest that different positions on the foundations of mathematics and in theories of mathematical practice can be seen as different positions on the admissibility of steps to each of these two elements. In the second paper, Manfred Kerber, Christoph Lange and Colin Rowat present their work on representing and proving theorems in economics, in particular their application of mechanised theorem proving tools to a class of economic problems for which very few general tools currently exist. Thirdly, Alison Pease and Ursula Martin use Turing's famous question "Can machines think?" as a starting point for a thought experiment about the question "Can machines do mathematics?", analysing an online mathematical discussion with a view to asking how machines could ever contribute to such a discussion. Alan Smaill argues, in the fourth paper, that the two distinctly different proposals for the choice of connectives in linear logic show the influence of different mathematical analogies at work. Finally, Sandra Visokolskis combines philosophy with the history of mathematics, giving an account of the sudden emergence of some mathematical results, characterising discovery in mathematics as an active and experiential process.

*Alison and Brendan*

July 2012

## **Programme Committee**

Bart van Kerkhove (Vrije Universiteit Brussel)

Brendan Larvor (University of Hertfordshire)

Tyler Marghetis (University of California at San Diego)

Alison Pease (Queen Mary University)

Dirk Schlimm (McGill University)

Alan Smaill (University of Edinburgh)

# Contents

- The Parallel Structure of Mathematical Reasoning* ..... 7  
Andrew Aberdein
- Formal Representation and Proof for Cooperative Games* ..... 15  
Manfred Kerber, Christoph Lange and Colin Rowat
- Seventy four minutes of mathematics: An analysis of the third Mini-Polymath project* ..... 19  
Alison Pease and Ursula Martin
- Mathematical Notation and Analogy* ..... 30  
Alan Smaill
- Discovery in Mathematics as an Experiential Practice of Privation* ..... 32  
Sandra Visokolskis

# The Parallel Structure of Mathematical Reasoning

Andrew Aberdein

**Abstract.** This paper proposes an account of mathematical reasoning as parallel in structure: the arguments which mathematicians use to persuade each other of their results comprise the argumentational structure; the inferential structure is composed of derivations which offer a formal counterpart to these arguments. Some conflicts about the foundations of mathematics correspond to disagreements over which steps should be admissible in the inferential structure. Similarly, disagreements over the admissibility of steps in the argumentational structure correspond to different views about mathematical practice. The latter steps may be analysed in terms of argumentation schemes. Three broad types of scheme are distinguished, a distinction which is then used to characterize and evaluate four contrasting approaches to mathematical practice.

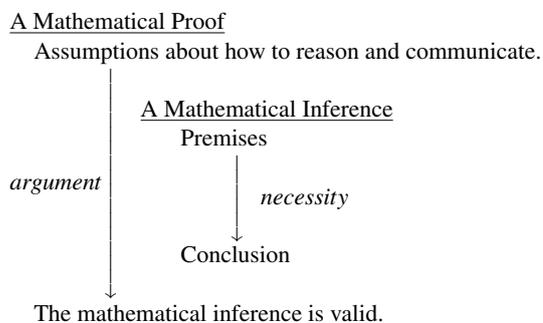
## 1 Structures of Mathematical Reasoning

Many mathematicians and philosophers of mathematics have observed the dual nature of mathematical proof: proofs must be both persuasive and rigorous. Here is G. H. Hardy:

If we were to push it to its extreme we should be led to a rather paradoxical conclusion; that we can, in the last analysis, do nothing but *point*; that proofs are what Littlewood and I call *gas*, rhetorical flourishes designed to affect psychology, pictures on the board in the lecture, devices to stimulate the imagination of pupils. . . . On the other hand it is not disputed that mathematics is full of proofs, of undeniable interest and importance, whose purpose is not in the least to secure conviction. Our interest in these proofs depends on their formal and aesthetic properties. Our object is *both* to exhibit the pattern and to obtain assent. [17, 18, his emphasis]

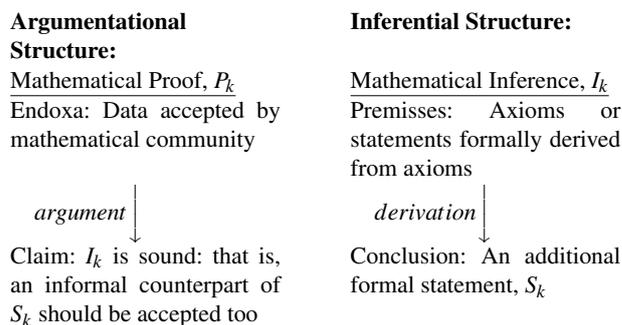
It follows from this account that ‘proof’ is ambiguous between two different activities: ‘exhibiting the pattern’ and ‘obtaining assent’. In most circumstances both activities must be satisfactorily performed for the proof to be a success. There are some special cases, such as proofs that have been fully formalized, or have been reified as mathematical objects, where only the first activity is attempted. But in the characteristic sense of ‘proof’ we need more than this; we need a dialectical interaction with the mathematical community.

For Richard Epstein, proofs intended to obtain assent are *arguments* by means of which mathematicians convince each other that the corresponding *inferences* are valid. He represents this situation schematically (Fig. 1). However, proofs are typically made up of many steps, not all of which are necessarily developed with the same rigour. So closer examination of proofs will represent them not as single arguments but as structures of arguments (technically trees, or directed acyclic graphs). Applying this detail to the broader picture suggested by Epstein requires the articulation of two parallel



**Figure 1.** Epstein’s picture of mathematical proof [15]

structures: an *inferential structure* of formal derivations linking formal statement to formal statement, and an *argumentational structure* of arguments by which mathematicians attempt to convince each other of the soundness of the inferential structure, and thereby of the acceptability of the informal counterparts of those statements. In Hardy’s terms, it is the inferential structure which is responsible for ‘exhibiting the pattern’ while the argumentational structure is responsible for ‘obtaining assent’. Fig. 2 summarizes this picture.



**Figure 2.** The parallel structure of mathematical reasoning

This account both conserves and transcends the conventional view of mathematical proof. The inferential structure is held to strict standards of rigour, without which the proof would not qualify as mathematical. However, the step-by-step compliance of the proof with these standards is itself a matter of argument, and susceptible to challenge. Hence much actual mathematical practice takes place in the argumentational structure. Careful demarcation of these two levels is essential to the proper understanding of mathematics. If this account is correct, important concepts in the philosophy of mathematics, such as mathematical rigour and mathematical explanation, can only properly be addressed when *both* of the parallel structures are accounted for. In order to do so, we need to say more about the details of the individual components of the two structures. The formal na-

ture of the inferential structure makes its characterization comparatively straightforward. It is a graph of vertices linked by edges, where the vertices are statements expressed in some formal system and the edges are derivations admissible in that system. The underlying formal system might, for example, be a natural deduction presentation of a particular system of logic, but more characteristically it will be a higher-level language. The argumentational structure poses more of a challenge, which I shall turn to in the next section.

## 2 The Argumentational Structure of Mathematics

Wilfred Hodges offers a brief analysis of ‘unformalised deductive argument[s]’ that is of help here (although I shall argue that it is necessary to generalize his account beyond deductive argumentation). For Hodges, such arguments contain components of three sorts:

- There are the stated conclusion, the stated or implied starting assumptions, and the intermediate propositions used in getting from the assumptions to the conclusion. I shall call these the *object sentences*.
- There are stated or implied justifications for putting the object sentences in the places where they appear. For example if the argument says ‘ $A$ , therefore  $B$ ’, the arguer is claiming that  $B$  follows from  $A$ .
- There are instructions to do certain things which are needed for the proof. Thus ‘Suppose  $C$ ’, ‘Draw the following picture, and consider the circles  $D$  and  $E$ ’, ‘Define  $F$  as follows’. [21, 6]

The argumentational structure must account for all three of these components.

The object sentences are the vertices of the argumentational structure. They are informal counterparts to formal statements of the inferential structure. Strictly speaking, the inferential structure must be grounded in axioms: propositions within the formal system whose truth must be granted if the system is to be employed. (As such the truth of the axioms cannot be established by formal proof; it is one area of mathematics where informal argument is the only option.) But practicing mathematicians seldom take things back that far. Instead, they begin as informal arguers do in any domain, with the informal counterpart of formal axioms, *endoxa*: ‘propositions considered true by everybody, or by the majority, or by the wise’ [42, 273]. In mathematics, the *endoxa* comprise what Reuben Hersh has called ‘established mathematics’:

Established mathematics is the body of mathematics that is accepted as the basis for mathematicians’ proofs. It includes proved statements ‘in the literature,’ and also some simpler statements that are so well accepted that no literature reference is expected. The central core of established mathematics includes not only arithmetic, geometry, linear and polynomial algebra, and calculus, but also the elements of function theory, group theory, topology, measure theory, Banach and Hilbert spaces, and differential equations—the usual first two years of graduate study. And then to create new mathematics, one must also master major segments of the established knowledge in some special area [19].

Thus, since informal mathematics need not always start with first principles, the argumentational structure will typically correspond only to a proper substructure of its inferential counterpart.

However, the correspondence will not be vertex to vertex: formal mathematics necessarily proceeds by very fine-grained increments, which the corresponding informal mathematical proofs typically elide. Thus many formal propositions of the inferential structure will have no counterpart object sentences in the argumentational structure. Conversely, some object sentences will have no counterpart formal propositions, since in some informal proofs there are object sentences representing intermediate statements which are not strictly needed for the validity of the proof, but only for its intelligibility [4].

As Hodges observes, the argumentational structure is more complex than the inferential structure because it contains instructions as well as justifications. The edges of the argumentational structure must be defined loosely enough to articulate both components. Hersh has the following to say about the linkages between object sentences:

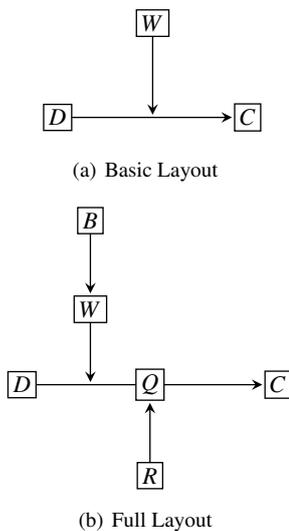
Established mathematics is an intricately interconnected web of mutually supporting concepts, which are connected both by plausible and by deductive reasoning. . . . Deductive proof, mutually supporting interconnections, and close interaction with social life (commerce, technology, education) all serve to warrant the assertions of established mathematics. Deductive proof is the principal and most important warrant [19].

Of course, this is still *unformalized* deductive proof, so it belongs in the argumentational structure not the inferential structure. Nonetheless, in many cases, the formalization is fully worked out, or more typically, it is clear how it could be. In this situation the corresponding steps of the argumentational structure can be very simple; they need do no more than point (as Hardy puts it) at the steps of the inferential structure.<sup>1</sup> But where the derivation is more complex or contested, much more of the burden of the proof rests on the argumentational structure. In those circumstances it becomes critical to track and provide responses to the objections that may be raised to the gaps in the inferential structure. As we shall see, exactly how this is to be achieved turns on what steps are admissible in the argumentational structure. Moreover, Hersh observes what I suggested above: even within established mathematics, not all the interconnections are deductive. Hence, some of the admissible steps will need to be framed in more permissive terms.

Further progress requires closer attention to the individual steps of the argumentational structure. One of the most influential attempts to analyze argumentational steps without appealing to logical form was developed in the 1950s by Stephen Toulmin. His ‘layout’ can represent deductive inference, but encompasses many other species of argument besides. In its simplest form, shown in Fig. 3(a), the layout represents the derivation of a Claim ( $C$ ), from Data ( $D$ ), in accordance with a Warrant ( $W$ ). This DWC pattern may appear to resemble a deductive inference rule, such as *modus ponens*, but it can be used to represent looser inferential steps.

The differences between the types of inference which the layout may represent are made explicit by the additional elements of the full layout shown in Fig. 3(b). The warrant is justified by its dependence on Backing ( $B$ ), possible exceptions or Rebuttals ( $R$ ) are allowed for, and the resultant force of the argument is stated in the Qualifier ( $Q$ ). Hence the full layout may be understood as ‘Given that  $D$ , we can  $Q$  claim that  $C$ , since  $W$  (on account of  $B$ ), unless  $R$ ’.

<sup>1</sup> Such ‘pointing’ may bring to mind Jody Azzouni’s ‘derivation indicator view’ of mathematical practice [8]. However, Azzouni’s ‘indicating’ describes a looser correspondence, closer to that holding in general between the two structures. Moreover, at least on some construals, such as that in [14, 149], Azzouni’s conception of derivation is broader than mine, so this is somewhat misleading.



**Figure 3.** Toulmin layouts

In a frequently cited example (derived from [40, 104]), ‘Given that HARRY WAS BORN IN BERMUDA, we can PRESUMABLY claim that HE IS BRITISH, since ANYONE BORN IN BERMUDA WILL GENERALLY BE BRITISH (on account of VARIOUS STATUTES . . .), unless HIS PARENTS WERE ALIENS, SAY.’ In recent years a number of authors have demonstrated that informal deductive proofs may be represented using Toulmin layouts (for example [5, 23]). Elsewhere I have shown how the layout may be generalized to exhibit arguments of a greater degree of structural complexity [1, 213 ff.].

The Toulmin layout is well-adapted to display the object sentences and justifications of specific steps of the argumentational structure. However, instructions, the last of Hodges’s three argument components would in general be difficult to fit into a Toulmin layout (although it has been argued that the warrant should be understood as an instruction not a proposition, [20, 71]). We also need a means to characterize generic steps, if structures are to be analyzed in terms of the admissibility of steps. One way of tackling both issues is in terms of *argumentation schemes*: stereotypical reasoning patterns, often accompanied by *critical questions*, which itemize possible lines of response. Schemes are framed in generic terms and there is no obstacle to their inclusion of instructions as well as object sentences and justifications (see [34, 49 f.] for an exploration of how this might be done). Like the Toulmin layout, schemes can be deductive, although they were devised to describe a more diverse range of arguments, such as argument from consequences or argument from expert opinion. (Chp. 9 of [42] catalogues sixty such schemes, many with multiple subvariants.) However, despite their topic-neutral provenance, many of these schemes are applicable to mathematical argumentation (as I have argued elsewhere: [2, 3]). Indeed, there is at least a family resemblance between argumentation schemes and Toulmin layouts. Most (instantiations of) schemes could be expressed as layouts: the data and warrant correspond to premisses, the claim (suitably modified by the qualifier) to the conclusion, and the backing and rebuttal both comprise possible answers to critical questions (for further dis-

ussion of the relationship of schemes to layouts, see [34, 28 ff.]).

We shall divide the schemes which will be needed to describe mathematical reasoning into three groups, on the basis of how their instantiations are related to the corresponding steps (if any) of the inferential structure. A-schemes correspond directly to a derivation rule of the inferential structure. (Equivalently, we could think in terms of a single A-scheme, the ‘pointing scheme’ which picks out a derivation whose premisses and conclusion are formal counterparts of its grounds and claim.) B-schemes are less directly tied to the inferential structure. Their instantiations correspond to substructures of derivations rather than individual derivations (and they may appeal to additional formally verified propositions). B-schemes may be thought of as exclusively mathematical arguments: high-level algorithms or macros that may in principle be formalized as multiple inferential steps. C-schemes are even looser in their relationship to the inferential structure, since the link between their grounds and claim need not be deductive. However, their instantiations may still correspond to substructures of the inferential structure, although there will be no guarantee that this is so and no procedure that will always yield the required structure even when it exists.

I shall make three initial observations about the relationships between the three types of scheme. Firstly, the classification is relative to the composition of the inferential structure. For example, if the inferential structure was a natural deduction presentation of classical logic, then only schemes corresponding to the rules of that system, such as modus ponens, would qualify as A-schemes; even fairly elementary mathematical deductive inferences would comprise B-schemes; and the C-schemes would cover inference patterns that were not in general deductive. Alternatively, if the inferential structure was a higher-level but constructive system, such as Nuprl [28, 285], many of these B-schemes would be A-schemes since they corresponded to higher-level rules, but some of the A-schemes, the constructively inadmissible ones, would now be C-schemes. Secondly, in terms of the Toulmin layout, the difference between the three types of scheme is tracked by their qualifier. The A-schemes would all be qualified as ‘by an immediate, formally valid derivation’, whereas the B-schemes would all be qualified as ‘by an in-principle formalizable valid deduction’. The C-schemes would have different qualifiers in different cases: typically something like ‘by plausible mathematical reasoning’, but sometimes weaker or stronger. This is a good reason to explicitly state the qualifier, at least of C-schemes, as Matthew Inglis has urged [23]. Thirdly, many schemes have subschemes: special cases, all of whose instantiations might also be seen as instantiations of the general scheme, but where additional constraints are imposed. But the classification of a scheme as belonging to a specific type need not transfer to its subschemes, which may well be more rigorous. For instance, the example mentioned above of a scheme containing instructions is a B-scheme, but also a subscheme of a C-scheme for analogy [34, 49 f.].

### 3 Four Views of Mathematical Practice

Different views about which steps should be admitted to the inferential structure give rise to different accounts of the foundations of mathematics. Most conspicuously, the divergence between classical and constructive mathematics may be characterized in these terms. Correspondingly, different views about which steps should be admitted to the argumentational structure give rise to different accounts of mathematical practice. In this fashion, at least four such accounts may be distinguished:

0. Only A-schemes are admissible. There is no such thing as ‘informal mathematical reasoning’: only formalized reasoning can count as mathematical. All that the argumentational structure can do is ‘point’ at the inferential structure.
1. Only A- and B-schemes are admissible. Informal mathematical reasoning is possible, but the argumentational structure must employ exclusively mathematical steps, albeit ones characterized informally.
2. All three types of scheme are admissible. Informal mathematical reasoning is possible, and the argumentational structure employs both exclusively mathematical steps, and steps of more general application.
3. Only topic-neutral A- or C-schemes are admissible. Informal mathematical reasoning is possible, and must be understandable purely in terms of steps of general application. No argumentational structure need contain any exclusively mathematical steps; that is, all such steps must be reducible to instances of general steps.

Support for each account may be found in mathematical practice. This section explores their competing merits.

### Option 0: No Such Thing as Informal Mathematics

The first question to ask about Option 0, the claim that all mathematics is formal, and hence that ‘informal mathematics’ is not mathematics, is whether it has ever been taken seriously. Both formalism and logicism require that mathematics be formalizable, but neither thesis need insist that mathematics isn’t mathematics until it has been formalized. This broader claim is more familiar as a polemical exaggeration, as with Russell’s insistence that the history of (pure) mathematics began in 1854 [37, 75], or as a straw man erected by critics of such programmes. Hence Lakatos informs us that ‘according to logical positivism, informal mathematics, being neither analytical nor empirical, must be meaningless gibberish’ ([25, 59]). While this may be a reasonable inference about the party line of a movement now rather lacking in adherents, specific endorsements are harder to find. For example, Carnap might seem a plausible candidate, but closer inspection reveals this is not actually so [27, 14]. Looking beyond logical positivism, even the arch-formalists of Bourbaki do not consistently practice what they preach, stressing instead the importance of ‘experience and mathematical flair’ [11, 8]. As the example of Bourbaki demonstrates, whether or not this position is credible as a regulative ideal, even its most single-minded proponents have difficulty living up to it.

A more fundamental defence of Option 0 (or perhaps a narrow interpretation of Option 1) arises from the wholesale rejection of informal reasoning proposed by David Miller. He states that ‘It cannot be denied that a complex sequence of interlocked blind guesses and cruel rejections may look much like directed thought, just as Darwinian evolution simulates orthogenesis or design. But we must not be hoodwinked into thinking that it is reasoning, or anything else that we know, that drives us forward to what is unknown. What reasoning does is pull us back’ [32, 68]. For Miller, the only legitimate task that arguments, deductive or otherwise, can perform is critical (and only deductive arguments are any use for this) [33, 65]. He canvasses, but rejects, three other possible roles: persuasion, discovery, and justification. Justification, he states, must entail regress or circularity, because an argument can only justify if its premisses are themselves justified. Persuasion he dismisses as unrelated to the argument: if the argument is thought persuasive, it can only be because the premisses are themselves persuasive, but then the argument makes no

contribution. His rejection of discovery as a property of argument is more tendentious: he regards any discoveries arising from valid inference as trivial, in so much as they would be already known to us, were we logically omniscient. But this leads to two important concessions: ‘To be sure, [deductive arguments] can provide new subjective knowledge, in the way that mathematical proofs uncannily do. Arguments, it may be conceded, do have an exploratory function, even if what they explore is what is already known, or conjectured, about the world, and not the world itself’ [32, 66]. Already known, that is, to the logically omniscient. Moreover, ‘inferences . . . that resist deductive reconstruction,’ while ‘evidently indistinguishable from blind guesses, . . . can indeed lead to an augmentation of knowledge (provided that knowledge is recognized, as it must be, to be conjectural through and through and through)’ [32, 67]. So, if our concern is with actual mathematical practice, in which logical omniscience is sadly unavailable, then informal deductive arguments, and even non-deductive arguments, can have a place in discovery.

In broader terms, we may observe that Miller’s challenge is most effective against a position dual to his own: that argument only draws us forward and never pulls us back. This would indeed be a dangerous position, especially if the steps of such arguments need not be deductive. Even if individual steps carry high levels of confidence, providing that the doubts are independent and thereby cumulative, multi-step arguments would swiftly grow less convincing as they lengthen. A system which encouraged us to pursue such arguments without correction could lead us far astray. But what of systems which permit both the forward propagation of confidence and the back propagation of doubt? Why must our guesses be blind? Certainly, prior experience can be a poor guide and even the most confident hunches can prove wide of the mark. But, if only for purposes of resource management, they are often the best place to start.

### Option 1: All Steps Exclusively Mathematical

B-schemes, the sort of schemes which are admissible at Option 1 but not at Option 0, comprise mathematical argument patterns of more than purely local application. For example, diagonalization, or finding the determinant of a matrix or the adjoint of an operator. In more complicated cases, multiple B-schemes can act in concert as a transferable technique which may be applied to diverse mathematical problems. This is what Jody Azzouni calls an *inference package*:

a capacity to recognize the implications of several assumptions by means of the representations of objects wherein those several assumptions have been knit together (psychologically). I also claim that the employment of inference packages shows up everywhere in mathematical practice. What an inference package allows a mathematician to do is to reason about a subject matter compatibly with a formal mechanical proof. His reasoning, however, is topic-specific and the various assumptions operative in that reasoning function together in a way that makes them phenomenologically invisible to him [9, 20].

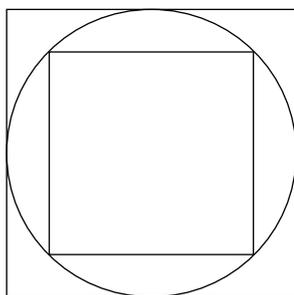
B-schemes are in principle formalizable, that is to say any instantiation of a given scheme in the argumentational structure should correspond to a substructure of the inferential structure in a predictable fashion. Nonetheless, actually working out what the latter structure should be in a given case may be forbiddingly laborious. Many B-schemes are intrinsic to established mathematics, and fluency in their use is a prerequisite for participation in mathematical practice. (For an attempt to construct an explicit B-scheme, see [34, 49 ff.].)

A lot of mathematics can be conducted using A- and B-schemes, so Option 1 is a much more reasonable characterization of mathematical practice than Option 0. However, as we shall see, it does have important limitations. To stick to Option 1, and thereby reject the admissibility of C-schemes, is to restrict mathematical practice to established mathematical techniques. Hence, Option 1 comprises a defence of the purity of proof method. This defence originates in Aristotle's rejection of 'metabasis', or 'kind crossing', the use of methods proper to one domain of reasoning within another. Indeed, some proponents of Option 1 appeal directly to Aristotle's arguments (for example, [24, 455 f.]). Thus, although geometrical methods would be permitted in astronomy, which Aristotle regards as a science subordinate to geometry, they should not be used in arithmetic [6, 75a]. Hence mathematics *are* genuinely plural: for Aristotle, they comprise several distinct domains whose methods overlap only by analogy [6, 76a]. In this respect he differs from Plato, for whom dialectic, that is informal reasoning, comprises a highest domain to which all others are subordinate [12, 179].

As an example of Aristotle's approach, consider the following:

Bryson's method of squaring the circle, even if the circle is thereby squared, is still sophistical because it does not conform to the subject in hand. So, then, any merely apparent reasoning about these things is a contentious argument, and any reasoning that merely appears to conform to the subject in hand, even though it be genuine reasoning, is a contentious argument: for it is merely apparent in its conformity to the subject matter, so that it is deceptive and plays foul [7, 171b].

Some unpacking is required. Sources both ancient and modern disagree as to precisely what Bryson's method comprised ([30, 18, 47 ff.]; [16, 35]). But there is consensus that it began by both inscribing and circumscribing the circle with squares, perhaps as in Fig. 4. One school of thought, attributed to Proclus, represents Bryson's argument as merely an existence claim: some square with an area between the two must be equal in area to the circle. Another, questionably attributed to Alexander of Aphrodisias, has Bryson proceed to inscribe and circumscribe pentagons, then hexagons, and so on. With each extra side the margin of error shrinks, so the method might charitably be interpreted as a partial anticipation of the method Archimedes used to estimate the area of a circle. However, in either case Aristotle would reject the result as fallacious. This has the uncomfortable corollary that, had he lived to see it, Aristotle would apparently have rejected Archimedes's method too.



**Figure 4.** Bryson's method of squaring the circle

Of course, in one sense, Aristotle would be right: Archimedes's method is not purely geometrical, and however far it is taken it does not constitute a solution to the ancient (and provably insoluble) problem of squaring the circle by a ruler and compass construction. The

difficulty with Aristotle's approach is whether he can acknowledge an innovative method as mathematical if it does not fit into any of the kinds of mathematics which he recognizes. This is exacerbated by his rejection of a common kind to which all other mathematical kinds would be subordinate. Later purists have been less reticent: Descartes proposed a *mathesis universalis* based on algebra; in the twentieth century set theory and subsequently category theory have been likewise proposed as common mathematical kinds. However, such a relationship between a general system and specific mathematical inferences is easier to reconcile with formal than with informal mathematics.

We may ask whether purity is an appropriate, or even intelligible ideal for informal mathematics. Methodological innovation often originates in informal mathematics. When a new method is first proposed, it can be controversial whether it qualifies as mathematical. Typically, its early use is restricted to the heuristic pursuit of results subsequently confirmed by more conventional means. Euler's employment of alternating series to establish that  $\sum_{n=1}^{\infty} 1/n^2 = \pi^2/6$  is one frequently discussed example [35]; the anticipation of the Laplace transform by Heaviside's operator methods another [22]. To insist on purity of method at this point may be appropriate if the informally derived results are (mis)represented as formally sound, but can otherwise only be an unnecessary brake on progress.

In conclusion, Aristotle may charitably be read as talking only of (*mutatis mutandis*) formal mathematics, and therefore not endorsing Option 1 after all. Latterday Aristoteleans, in so far as they insist on 'certain conventions to which an argument must conform to be an argument within the discipline' [24, 455 f.], beg the question against methodological innovation. Requiring that the new method be shown to be mathematical before it can be admitted into informal mathematical reasoning is to require that it be shown to be either formally valid or heuristically useful. But the former is inappropriate for informal mathematics and the latter can only be demonstrated through extensive use.

## Option 2: Not All Steps Exclusively Mathematical

The Victorian mathematician J. J. Sylvester, in defending his discipline from the ill-informed criticism of T. H. Huxley, stressed 'how much observation, divination, induction, experimental trial, and verification, causation, too (if that means, as I suppose it must, mounting from phenomena to their reasons or causes of being<sup>2</sup>) have to do with the work of the mathematician' [39, 1762]. This perspective emphasizes the connexions between mathematical practice and ordinary reasoning, much of it comprising C-schemes, thereby leading to Option 2 or, when the ordinary reasoning is understood as underpinning the mathematical practice, Option 3.

One important challenge to Option 2 arises from an ambiguity between two senses of the phrase 'mathematical argument'. The ambiguity may be resolved by applying the terminology of Toulmin layouts. When 'mathematical argument' occurs in ordinary, that is non-mathematical, discourse it often serves to indicate that the argument has a mathematical warrant and/or backing, and thereby a mathematical qualifier. For example, in a discussion on the prospects of two competing political candidates it might be argued that one candidate is too far behind to win, even if all the uncounted ballots had been cast in his favour. The proponent of such an argument may stress that it was a 'purely mathematical argument', thereby emphasizing its difference from typical political arguments: a more robust

<sup>2</sup> This would seem to be an attempt to characterize what would later be described as abductive reasoning, or inference to the best explanation.

warrant and qualifier. If that were the only context in which ‘mathematical argument’ were used it would appear that the mathematical nature of the warrant (or qualifier) was essential, thereby ruling out topic-neutral schemes, and thereby Option 2. But this would be to ignore the non-mathematical context—the warrant is singled out since it is what makes the argument different from other arguments to the same claim. In broader terms, we might also classify this as a political argument, since it derives a political claim from political data. In this second sense, a mathematical argument will be one which arises in mathematical discourse. Characteristically, this will entail the data and claim being mathematical, but not necessarily that the warrant should be so too.

If warrants in mathematical arguments are not mathematical what are they? The positive argument in defence of the mathematical use of C-schemes can best be made by answering this question. As discussed in Sect 2, although argumentation schemes have been developed primarily for non-mathematical contexts, many of them lend themselves readily to mathematical application. For example, consider the following scheme for argument from analogy:

ARGUMENTATION SCHEME FOR ARGUMENT FROM ANALOGY

**Similarity Premise** Generally, case  $C_1$  is similar to case  $C_2$ .

**Base Premise**  $A$  is true (false) in case  $C_1$ .

**Conclusion**  $A$  is true (false) in case  $C_2$ .

Critical Questions:

**CQ1.** Are there differences between  $C_1$  and  $C_2$  that would tend to undermine the force of the similarity cited?

**CQ2.** Is  $A$  true (false) in  $C_1$ ?

**CQ3.** Is there some other case  $C_3$  that is also similar to  $C_1$ , but in which  $A$  is false (true)? [42, 315]

This is a C-scheme because it is topic-neutral and not necessarily deductive. Whether or not this scheme is deductively valid turns on the interpretation of the warrant, that is the similarity premise. If ‘similar’ is defined narrowly enough, providing an equivalence between ostensibly unrelated mathematical structures, the premise would be categorical, and the scheme valid. This can be a source of formidable insight. The graph theorist Bill Tutte gives an impressive example in describing how he and three colleagues came to falsify Lusin’s Conjecture, that no square can be dissected into smaller squares, all of different sizes: ‘Eventually we found that a squared rectangle was equivalent to an electrical network of unit resistances. . . . We now understood that the basic theory of squared rectangles was that of Kirchoff’s Laws of electrical networks. That was something we could look up in textbooks’ [41, 3 ff.]. The similarity premiss is the equivalence that Tutte and his colleagues found, which could be formally demonstrated. Hence the base premisses, which they looked up in textbooks, were known to have true analogues in the theory of squared rectangles. This suggests that this specific case of analogy may also instantiate some B-scheme which is a subscheme of the general C-scheme for analogy above. One such subscheme is discussed in [34, 49 f.], but a better fit may be found in the account of analogy in [10], which may be construed as the articulation of several such subschemes.

But even if the similarity premise is defeasible, the scheme characterizes a versatile pattern-spotting procedure which can be an invaluable source of mathematical hypotheses. (This is not to prejudice the question whether defeasible analogies are *only* useful as heuristics.) A familiar example is the extension of the number concept,

from natural numbers to rational numbers, irrational numbers, negative numbers, imaginary numbers, transfinite numbers, and so forth. Each of these steps proceeded by analogy, and was initially controversial. These moves also gave rise to further analogies, as properties known to apply to one sort of number were conjectured to apply to other sorts. For example, Wallis conjectured that  $n^p \times n^q = n^{p+q}$ , known to apply for integer  $p, q$ , also held when the indices were fractional or negative [38, 158].

Informal analogies can be drawn from areas outside mathematics. For example, the optimization technique of ‘simulated annealing’ is motivated by an analogy with the metallurgical technique of annealing [31, 120]. The mathematical problem is that many search procedures designed to find the global maximum of functions of multiple independent variables will halt at local optima, that is globally suboptimal solutions. The analogous problem in metallurgy is that metals (and other crystalline substances) if cooled too quickly from a molten state will ‘freeze in’ irregularities, producing a structure of higher energy level, and typically weaker structural integrity. The solution is gradual cooling, which finds an analogy in the incremental reduction of a control parameter in the simulated annealing algorithm. Here, the base premiss, that is the data, is drawn from metallurgy, not mathematics, while the warrant asserts the analogy between the two fields, and is thus not wholly mathematical either. Nonetheless, the claim is a substantial contribution to a mathematical problem. While more formal arguments have been made for the effectiveness of simulated annealing, as they must be if it is to have a permanent place in mathematics, this analogy is independently compelling, and was crucial to its discovery.

We have seen that the data, the warrant, and thereby the qualifier of a mathematical argument may be non-mathematical. However, arguments with non-mathematical claims would *ipso facto* not be contributions to mathematics, but rather mathematical contributions to whatever discourse the claim was drawn from. What of the other components of the Toulmin layout? The susceptibility of mathematical argument to what Lakatos describes as ‘heuristic falsifiers’ would suggest the admissibility of non-mathematical rebuttals [26, 36]. (If a heuristic falsifier succeeds, the inferential structure is unaffected, since it is still formally valid, but the argumentational structure would be forced to follow a different path, bypassing the ‘falsified’ content.) And even ostensibly mathematical warrants may have non-mathematical backing, as the next section demonstrates.

### Option 3: No Steps Exclusively Mathematical

The distinction between Option 2 and Option 3 turns on whether all the instantiations of mathematical schemes in the argumentational structure are reducible to instantiations of non-mathematical schemes, or more precisely, whether all instantiations of B-schemes are decomposable into sequences or combinations of instantiations of A- and C-schemes. Advocates of Option 2 hold that room must be left for the irreducibly mathematical, even in the argumentational structure; proponents of Option 3 disagree.

This dispute may be difficult to resolve. On the one hand, there are elementary topic-neutral A- and C-schemes in mathematics, and many instantiations of more complex B-schemes can be reduced to instantiations of such schemes in an obvious enough if tedious fashion. Such, after all, is one of the goals of proof. On the other hand, there are many B-schemes of formidable complexity, for example those invoking widely used but hard to prove results. Even if B-schemes of this character can be understood as ultimately depending on elementary, topic-neutral A- or C-schemes, such schemes would

seem so remote from the use to which the B-scheme was put, especially in informal contexts, as to be of no practical relevance. But notice that this ostensible counterexample to Option 3 makes a crucial concession: if the arguer is merely using the result, and not reestablishing it on the fly, then the backing for his argument is not wholly mathematical. Rather he is ultimately relying, quite properly, on an argument from expert opinion. If the argument is challenged, the arguer will typically make the backing more explicit to show how the contentious scheme is supported, thereby answering one of its critical questions. This may take the form of restating the argument in more elementary steps, thereby bringing it closer to topic-neutral schemes, or the arguer may be forced to acknowledge that he is using a result which he is unable to prove, at least on demand, but has on good authority, at which point he has clearly resorted to a non-mathematical scheme.

More direct motivation for Option 3 may be found in the widely held thesis that ‘there are certain basic forms of thought and argument which are prior to the development of formal Mathematics’ [36, 237]. Such intuitions have taken a variety of forms. For example, Jody Azzouni once developed (but subsequently repudiated) a view of informal mathematics as a subdoxastic process analogous to that governing grammar, from which he infers that ‘all reasoning is topic-neutral in nature’ [9, 18 f.]. Keith Devlin reaches a similar conclusion from the grounds that mathematical ‘thought processes are comprised of brain activation patterns that are associated with real world stimuli’ [13, 378]. Penelope Maddy proposes basing mathematics on ‘a characterization of classical logic as grounded in a rudimentary logic that’s both true of the world and embedded in our most primitive modes of cognition’ [29, 288]. As Ian Dove concludes, ‘mathematical reasoning is already in accord with principles and techniques from informal logic—even if this is unnoticed by the practitioners’ [14, 150].

We have seen that all four options have defenders amongst both mathematicians and philosophers. However, Option 0 relies on either an untenable idealization of mathematical practice, or an arbitrary restriction of ‘mathematics’ to exclude much of that practice. And Option 1 is hard to reconcile with methodological innovation. That leaves Option 2 and Option 3, both of which leave room for informal reasoning in mathematical practice. Resolving the debate between these positions turns on whether all inferences of informal mathematics may be reduced to combinations of topic-neutral inferences. A final answer to this question is beyond the scope of this paper, but I would suggest that the onus is on proponents of Option 2 to provide examples of informal mathematical inferences for which no such reduction is possible.

## 4 Conclusion

This paper has defended an account of mathematical reasoning as comprised of two parallel structures. The argumentational structure is composed of arguments by means of which mathematicians seek to persuade each other of their results. The inferential structure is composed of derivations which offer a formal counterpart to these arguments. The precise relationship between the two structures may be understood in terms of the range of argumentation schemes which may be instantiated by steps of the argumentational structure. Just as different views about the foundations of mathematics may be characterized in terms of the admissibility of steps in the inferential structure, different views about mathematical practice may be characterized in terms of the admissibility of steps in the argumentational structure. I have made the case that a wide range of schemes

should be admitted to the argumentational structure. Two distinct areas emerge for further exploration: firstly, the investigation of fine-tuned, specifically mathematical B-schemes; and secondly, the application to mathematics of topic-neutral C-schemes.

## REFERENCES

- [1] A. Aberdein, ‘Managing informal mathematical knowledge: Techniques from informal logic’, in *MKM 2006*, eds., J. M. Borwein and W. M. Farmer, number 4108 in LNAI, pp. 208–221, Berlin, (2006). Springer-Verlag.
- [2] Andrew Aberdein, ‘Observations on sick mathematics’, in *Philosophical Perspectives on Mathematical Practice*, eds., Bart Van Kerkhove, Jean Paul Van Bendegem, and Jonas De Vuyst, 269–300, College Publications, London, (2010).
- [3] Andrew Aberdein, ‘Rationale of the mathematical joke’, in *Proceedings of AISB 2010 Symposium on Mathematical Practice and Cognition*, eds., Alison Pease, Markus Guhe, and Alan Smaill, 1–6, AISB, Leicester, (2010).
- [4] Jesse Alama and Reinhard Kahle, ‘Checking proofs’. Unpublished, 2012.
- [5] Jesús Alcolea Banegas, ‘L’argumentació en matemàtiques’, in *XIIIè Congrés Valencià de Filosofia*, ed., E. Casaban i Moya, 135–147, València, (1998). Translation online at <http://my.fit.edu/%TEaberdein/Alcolea.pdf>.
- [6] Aristotle, ‘Posterior analytics. Translated by G. R. G. Mure’, in *Introduction to Aristotle*, ed., Richard McKeon, 9–109, Random House, New York, NY, (1947).
- [7] Aristotle, ‘On sophistical refutations. Translated by W. A. Pickard-Cambridge’, in *Fallacies: Classical and Contemporary Readings*, eds., Hans V. Hansen and Robert C. Pinto, 19–38, Pennsylvania State University Press, University Park, PA, (1995).
- [8] Jody Azzouni, ‘The derivation-indicator view of mathematical practice’, *Philosophia Mathematica*, **12**(2), 81–105, (2004).
- [9] Jody Azzouni, ‘Why do informal proofs conform to formal norms?’, *Foundations of Science*, **14**(1–2), 9–26, (2009).
- [10] Paul Bartha, *By Parallel Reasoning: The Construction and Evaluation of Analogical Arguments*, Oxford University Press, New York, NY, 2010.
- [11] Nicolas Bourbaki, *Elements of Mathematics: Theory of Sets*, Springer, Berlin, 1968.
- [12] Mic Detlefsen, ‘Purity as an ideal of proof’, in *The Philosophy of Mathematical Practice*, ed., Paolo Mancosu, 179–197, Oxford University Press, Oxford, (2008).
- [13] Keith Devlin, ‘A mathematician reflects on the useful and reliable illusion of reality in mathematics’, *Erkenntnis*, **68**(3), 359–379, (2008).
- [14] Ian J. Dove, ‘Towards a theory of mathematical argument’, *Foundations of Science*, **14**(1–2), 137–152, (2009).
- [15] Richard L. Epstein, ‘Mathematics as the art of abstraction’. Unpublished, 2012.
- [16] R. J. Hankinson, ‘Aristotle on kind-crossing’, in *Philosophy and the Sciences in Antiquity*, ed., R.W. Sharples, 23–54, Ashgate, Aldershot, (2005).
- [17] G. H. Hardy, ‘Mathematical proof’, *Mind*, **38**, 11–25, (1928).
- [18] T. L. Heath, *Mathematics in Aristotle*, Clarendon, Oxford, 1949.
- [19] Reuben Hersh, ‘To establish new mathematics, we use our mental models and build on established mathematics’, in *From an heuristic point of view: In honor of Carlo Cellucci*, Forthcoming, (2012).
- [20] David Hitchcock, ‘Toulmin’s warrants’, in *Anyone Who Has a View: Theoretical Contributions to the Study of Argumentation*, eds., F. H. van Eemeren, J. Blair, C. Willard, and A. F. Snoeck-Henkemans, 69–82, Kluwer, Dordrecht, (2003).
- [21] Wilfrid Hodges, ‘An editor recalls some hopeless papers’, *Bulletin of Symbolic Logic*, **4**(1), 1–16, (1998).
- [22] Bruce J. Hunt, ‘Rigorous discipline: Oliver Heaviside versus the mathematicians’, in *The Literary Structure of Scientific Argument*, ed., Peter Dear, University of Pennsylvania Press, Philadelphia, PA, (1991).
- [23] Matthew Inglis, Juan Pablo Mejía-Ramos, and Adrian Simpson, ‘Modelling mathematical argumentation: The importance of qualification’, *Educational Studies in Mathematics*, **66**(1), 3–21, (2007).
- [24] Erik C. W. Krabbe, ‘Strategic maneuvering in mathematical proofs’, *Argumentation*, **22**(3), 453–468, (2008).

- [25] Imre Lakatos, 'Cauchy and the continuum: The significance of non-standard analysis for the history and philosophy of mathematics', in *Philosophical Papers*, volume 2, 43–60, Cambridge University Press, Cambridge, (1978).
- [26] Imre Lakatos, 'A renaissance of empiricism in the recent philosophy of mathematics', in *Philosophical Papers*, volume 2, 24–42, Cambridge University Press, Cambridge, (1978).
- [27] Gregory Lavers, 'Carnap, formalism, and informal rigour', *Philosophia Mathematica*, **16**(1), 4–24, (2008).
- [28] Donald MacKenzie, *Mechanizing Proof: Computing, Risk, and Trust*, MIT Press, Cambridge, MA, 2001.
- [29] Penelope Maddy, *Second Philosophy: A Naturalistic Method*, Oxford University Press, Oxford, 2007.
- [30] Henry Mendell, 'Bryson's squaring of the circle'. <http://www.calstatela.edu/faculty/hmendel/Ancient%20Mathematics/Philosophical%20Texts/Bryson/Bryson.html>, n. d.
- [31] Zbigniew Michalewicz and David B. Fogel, *How to Solve It: Modern Heuristics*, Springer, Berlin, 2004.
- [32] David Miller, 'Do we reason when we think we reason, or do we think?', *Learning for Democracy*, **1**(3), 57–71, (2005).
- [33] David Miller, *Out of Error: Further Essays on Critical Rationalism*, Ashgate, Aldershot, 2006.
- [34] Alison Pease and Andrew Aberdein, 'Five theories of reasoning: Interconnections and applications to mathematics', *Logic and Logical Philosophy*, **20**(1-2), 7–57, (2011).
- [35] George Pólya, *Mathematics and Plausible Reasoning*, Two Volumes, Princeton University Press, Princeton, N.J., 1954.
- [36] Abraham Robinson, 'Formalism 64', in *Proceedings of the International Congress for Logic, Methodology and Philosophy of Science, Jerusalem*, ed., Y. Bar-Hillel, 228–246, North-Holland, Amsterdam, (1964).
- [37] Bertrand Russell, 'Mathematics and the metaphysicians', in *Mysticism and Logic*, 75–95, Unwin, London, (1986 [1901]).
- [38] Jacqueline A. Stedall, *A Discourse Concerning Algebra: English Algebra to 1685*, Oxford University Press, Oxford, 2002.
- [39] James Joseph Sylvester, 'The study that knows nothing of observation', in *The World of Mathematics*, ed., James R. Newman, volume 3, 1758–1766, Simon & Schuster, New York, NY, (1956 [1869]).
- [40] Stephen Toulmin, *The Uses of Argument*, Cambridge University Press, Cambridge, 1958.
- [41] W. T. Tutte, *Graph Theory As I Have Known It*, Oxford University Press, Oxford, 1998.
- [42] Douglas N. Walton, Chris Reed, and Fabrizio Macagno, *Argumentation Schemes*, Cambridge University Press, Cambridge, 2008.

# Formal Representation and Proof for Cooperative Games

Manfred Kerber and Christoph Lange and Colin Rowat<sup>1</sup>

**Abstract.** In this contribution we present some work we have been doing in representing and proving theorems from the area of economics, and mainly we present work we will do in a project in which we will apply mechanised theorem proving tools to a class of economic problems for which very few general tools currently exist. For mechanised theorem proving, the research introduces the field to a new application domain with a large user base; more specifically, the researchers are collaborating with developers working on state-of-the-art theorem provers. For economics, the research will provide tools for handling a hard class of problems; more generally, as the first application of mechanised theorem proving to centrally involve economic theorists, it aims to properly introduce mechanised theorem proving techniques to the discipline.

## 1 Introduction

*Finding stable sets [in cooperative games] involves a new tour de force of mathematical reasoning for each game or class of games that is considered. Other than a small number of elementary truisms ... there is no theory, no tools, certainly no algorithms.* R. Aumann (Economics Nobel laureate of 2005), 1987 [1].

While proving mathematical theorems is often seen as a rigorous process from the outside, it remains, as a form of human reasoning, an error-prone process (e.g. through the omission of hidden assumptions). It is also costly. A system must first be formalised by selection of an axiom set. An appropriate solution concept must then be posited. Third, its properties must be derived, typically existence of the solution, uniqueness, and perhaps smoothness. Finally, if the solution does not behave as expected, the axioms may be adjusted, and the process be restarted. The third step, and its repetitions, is typically the most expensive, and certainly error-prone. This is particularly true when working with new axiomatic systems for which researchers have only limited intuitions.

Lakatos' classic study of this process used Euler's polyhedron formula [20], but examples can be found in any discipline. In cooperative game theory – the application domain of the research presented here – even the most respected researchers illustrate it. For instance, in [24], which laid the foundations of modern game theory, von Neumann and Morgenstern assumed that their solution concept, the stable set, always existed in the most widely studied class of cooperative games; this assumption was later shown false [21]. More recently, Nobel laureate Maskin [22] claimed in a presidential address to the Econometric Society, economics' leading society for economic theory, that certain properties of a game generalised beyond  $n = 3$ ; counterexamples to this claim were found [7].

Errors and confusion are undesirable for purely intellectual reasons. Further, when the results claimed may come to inform public policy, as can happen with applications of economic theory to the financial world, additional concerns arise and the consequences can be severe. For both of these reasons it is desirable to make use of the recent progress both in mechanised theorem proving and in formally supporting human researchers. Mechanised theorem proving systems have reached a level which allows their application to previously intractable areas; proof development environments are successfully used in areas as diverse as hardware verification [12], software verification (in particular in safety relevant areas [27]), mathematical proof development (e.g., the advanced research mathematics in Hales' Flyspeck project [11]), and educational systems [23]. All security sensitive software must satisfy high standards of formal verification (Evaluation Assurance Level 7, the highest standard [6]) to be approved by national agencies or accredited evaluation facilities. In addition to providing valuable services to new application areas, these developments have also led to insights into the strengths and weaknesses of the systems themselves.

In addition to reducing the possibility of error or confusion, formal methods also enable the reuse of proof tactics and offer the eventual promise of making the process of theory development and testing faster and less costly.

## 2 Degrees of formalisation

This research uses different types of mathematical knowledge representation and theorem proving tools.

**Semi-formal representation systems** The mathematical knowledge incorporated in mathematical theorems is typically given in the form of a PDF document, or a  $\LaTeX$  source document. Formal representation systems such as  $s\TeX$  [19], a semantic extension for  $\LaTeX$  make it possible to not only generate high-quality PDF output, but also to disambiguate mathematical text and thus make it accessible to further semantic services, which become available after translation to the more machine-friendly XML-based OMDoc format [18]: semantic content search as well as soft checks of the soundness of mathematical work. Such services demonstrate their full potential when they are applied to a whole *collection* of mathematical documents published on the Web (possibly even interlinked with external, related collections); semantic publication formats such as HTML 5 allow readers to interact with these services from within documents published in a human-readable way [8]. A system like  $s\TeX$  can therefore be the starting point of semi-formal services;  $s\TeX$  and OMDoc have been successfully applied to the representation of mathematics, software engineering [17], physics [13] and other mathematics related fields. Economics will be an exciting new application area.

---

<sup>1</sup> University of Birmingham, Birmingham, England

**Theorem provers** Theorem provers allow mathematical theorems to be formally proved, either fully automatically (q.v. systems such as Theorema or Leo), or interactively by allowing the user to develop a machine checked proof (q.v. Isabelle). Systems differ in a number of ways, including whether they use first order or higher order logic, types or not. This has consequences on the difficulty of beginning to work with them. In Theorema, an untyped system, an initial formalisation can be done relatively painlessly for users familiar with the Mathematica computer algebra system, in which Theorema is implemented and whose representation philosophy it follows. In Isabelle and Leo, both typed systems, this initial step is much harder; however, some mistakes, inaccuracies, and inconsistencies are discovered by the type system. Proving theorems is, however, generally difficult and, except for relatively easy theorems, usually requires at least some interactivity (for automated theorem provers the interaction means typically the adjustment of search options, reformulations of the problem, and the introduction of lemmas).

The usage of new theorem proving tools poses problems even for users with experience in the field. Some typical examples of such problems are:

- How to exactly formulate a particular concept and to be consistent with the formulation.
- To know whether the formulation chosen is adequate and the prover should be able to prove the theorems in this form or not.
- Which system to use for a particular problem.

Ideally there would exist a rulebook or cookbook when to use which system and how to apply it. While there are certain rules of thumb on when and how to apply systems – such as for a first order problem use a first order theorem prover and not a higher order one, likewise for a propositional logic problem a SAT solver – they are not strong enough to guide inexperienced users. A rule of thumb should also not be taken too seriously, since the same problem (e.g., Arrow’s theorem), may be formulated and proved in different logics. Systems may solve problems which are very hard for humans since they can search big search spaces very efficiently, but may not solve other problems with which humans have almost no problems. The latter may be the case since humans are often guided by an intuition – based on experience, examples, counterexamples – which is typically hard to add to an automated system. As a consequence humans often have to give a lot of detailed information about a proof, which would be left implicit on paper, in order to make it formal.

It should be noted that even if statements are not fully proved formally a formalisation can result in a much higher degree of accuracy. As soon as you start working with a formalisation – e.g., by stating some simple consequences – you may detect some unwarranted consequences which should lead to an improved formalisation.

The problems generated in our research will form a new class of challenge problems for automated theorem provers. The theorems in our application domain are typically formulated in higher order logic, but it is not clear whether all (or which ones) are inherently higher order. Correspondingly it will be investigated how many and which theorems can be proved automatically by first order theorem provers such as Vampire, or whether a higher order theorem prover such as Leo [2] can better deal with them. It will also be interesting to see how these systems compare with Theorema as an untyped system. Theorema has the additional advantage to allow for mixing reasoning and computation. The problem formalisations will be exported to the TPTP archive [28] as a new class of challenge problems for automated theorem provers.

### 3 Mechanised reasoning within economics

Modern economics has been transformed by a range of reasoning support tools, including computer algebra systems (e.g. Maple and Mathematica), statistical packages (e.g. STATA and R) and game solvers (e.g. Gambit). However, the advances in formalised mathematics and mechanised reasoning have yet to have had a discernible impact on economics. Mathematical knowledge representation and mechanised reasoning has been applied to known results in economics: results in social choice have attracted the most attention, including studies by Wiedijk [30] using Mizar, Nipkow [25] using HOL, and Grandi and Endriss [10] using Prover9. Non-cooperative game theory has also received some attention, including by Vestergaard [29] (with Coq). None of this research has involved economic theorists, nor has it been published in economics journals, giving it a marginal effect on economics to date. Thus, given the clear importance of these techniques, leading economic theorists have maintained a ‘watching brief’, aware that they will eventually be adopted by the discipline, but unsure as to when or how.

### 4 Cooperative game theory

Game theory models agents’ strategic interactions. The more familiar branch of the field is non-cooperative game theory, which defines a game form (a set of permissible moves) and then seeks a Nash equilibrium (a strategy for each agent such that none can do better by unilaterally selecting a different strategy). Cooperative game theory, on the other hand, abstracts from specific details of play, instead defining an irreflexive binary relation, called dominance, directly on the set of outcomes. In complex social environments, cooperative game theory may yield more robust insights than those tied to a particular game form.

Solutions to cooperative games are sets of outcomes satisfying conditions on dominance, principally the core (the set of undominated allocations) and a stable set (a set in which no element dominates another element in the set, but each other element in the whole space is dominated by an element in the stable set), von Neumann and Morgenstern’s original solution concept. While the core is a computationally simple structure, requiring no worse than pairwise tests of the dominance relation, the stable set is not. Even in certain simple games, the question of whether an outcome belongs to a stable set is NP-complete [4]; more generally, Deng and Papadimitriou [9] have argued that the question of whether a stable set even exists can be undecidable.

We work with pillage games, a class of cooperative games introduced by Jordan [14] and studied by us since 2007. Pillage games are defined by imposing monotonicity on the dominance relation to capture two basic intuitions: coalitions gain power as they gain members, and as their members gain resources. Technically, this monotonicity allows the dominance relation to be represented by a power function. Practically, pillage games help model contests of power – quintessentially complex social situations (e.g. political parties’ redistribution of resources to benefit their supporters). To our knowledge, our work represents the first attempt to formally prove properties of a cooperative game.

Pillage games are well-suited for our purposes for a number of reasons. First, we know that they are tractable. Second, as they have only been introduced recently, it is possible to formally represent everything known about pillage games in a reasonable amount of time. Third, they form a richer class of games than the two most commonly studied classes of cooperative games. Fourth, the formal similarities

between cooperative games and graphs (q.v. [3]) may allow rapid application of results to network economics, operations research, and combinatorial optimisation (including matching problems).

## 5 Work Plan

We will formally represent pillage games, and formally prove their key properties. A semi-formal representation will be done in sTeX. This is relatively inexpensive and allows efficient search for properties (such as the respective dependencies of definitions, lemmas, and theorems), aiding the reuse of results in the development of new theories.

Formal proofs will be generated using Theorema and Isabelle, and possibly a third system (likely from Wiedijk’s list at [31]), allowing comparison of the systems. This will result in great reliability of the theorems proved. As the theorems are typically of a higher order logic type, we will generate challenge problems for higher order theorem provers and integrate them into the TPTP problem library, fully automated theorem provers such as Leo will be tested on these problems. Proof tactics will be generated to simplify the generation of formal proofs.

Using the formalisation developed, new results on pillage games will be sought by relaxing some of the additional axioms imposed on top of pillage games’ basic monotonicity axioms. We shall primarily aim to remove the symmetry axiom. By allowing consideration of non-identical agents, this would ease attempts to empirically falsify theoretical predictions. Efforts at falsification are particularly useful in cooperative game theory, which has almost no associated empirical or experimental literature. Newly developed knowledge will be formally represented and investigated to ascertain which proof tactics used can be re-used (or adjusted).

Mechanised theorem proving may also provide important metaphors for economic research into bounded rationality, which has tended to use the number of cells in a finite automaton as its complexity measure. Mechanised theorem proving offers other measures, including search depth, tree-span, and perhaps most interestingly, the low-cost encoding of repeatedly used strategies as tactics.

Mechanised theorem proving as a field will benefit from a new and challenging application domain with a large user base. Testing theorem provers on a class of problems for which they have not been designed may highlight shortcomings, and suggest developments in theorem proving. The experience gathered in this project will directly feed in the reimplementations of the Theorema system.

For economic theory, the research will provide a proof of concept, providing tools and algorithms for handling a famously hard class of problems. It will provide a focal point for putting proofs in theoretical economics to a new level of reliability. This will lead to an improved reliability of the results in a field which studies complex social systems on which we all rely.

## 6 Illustrative example

To illustrate what is proposed, consider the first lemma from [15].

**Lemma 1** *Any power function  $\pi(C, \mathbf{x})$  can be represented by another,  $\pi'(C, (x_i)_{i \in C})$ , which depends only on the resource holdings of its coalition members.*

We have formalised and proved this lemma (and some others) in the Isabelle theorem proving system [26] as well as in the Theorema system [5], see [16].

The Theorema formalisation is:

**Lemma** [“powerfunction-independent”, any[ $\pi, n, C, x, y$ ], with[allocation<sub>n</sub>[ $x$ ]  $\wedge$  allocation<sub>n</sub>[ $y$ ]  $\wedge$   $C \subseteq I[n] \wedge$  powerfunction[ $\pi, n$ ],  $\forall_{i \in C} (x_i = y_i) \implies (\pi[C, x] = \pi[C, y])$  ]

The proof uses the weak resource monotonicity axiom of pillage games, **WR**: “if  $y_i \geq x_i \forall i \in C \subseteq I$  then  $\pi(C, \mathbf{y}) \geq \pi(C, \mathbf{x})$ ” and basic properties of real numbers, in particular, the law of trichotomy that for any two real numbers  $a$  and  $b$  holds one of the three cases  $a < b$ ,  $a = b$ , or  $a > b$ .

**Proof:** Consider arbitrary  $\mathbf{x}, \mathbf{y}$  such that  $x_i = y_i \forall i \in C \subseteq I$ . Then  $y_i \geq x_i$  and  $x_i \geq y_i$  so that axiom **WR** requires  $\pi(C, \mathbf{y}) \geq \pi(C, \mathbf{x}) \geq \pi(C, \mathbf{y})$ . For this to hold,  $\pi(C, \mathbf{x})$  cannot depend on  $x_j$  for any  $j \notin C$ .

The concepts involved in the lemma are mathematically relatively simple and easy to prove (others are considerably more difficult). However, a logical formulation can be more complex than expected and several design decisions have to be taken. Considerable efforts will be put in the question on how to facilitate an easy start for users of systems such as Isabelle and Theorema, who are not experts in theorem proving. The initial hurdle for the typed system of Isabelle is higher than that for the untyped Theorema system. However, it is much easier to get things wrong in an untyped system. Concretely, the first concept – that of a power function – was the most difficult one, since it was necessary to represent a function with range of a subset of agents and an  $n$ -tuple of real numbers adding up to one. Once this concept was represented other concepts were relatively standard to represent.

## 7 Conclusion

We have started to formalise a (small) area of theoretical economics in different ways and to different degrees in order to be able to study the benefit that can be obtained from these formalisations. We are at the start of the project and would like to get feedback from the participants of the symposium also to direct us to promising approaches and to learn about related work.

## REFERENCES

- [1] R. J. Aumann. What is game theory trying to accomplish? in: K.J. Arrow, S. Honkapohja (Eds.), *Frontiers of Economics*, Blackwell, Oxford, 1987.
- [2] C. Benzmüller et al. Progress report on Leo-II, an automatic theorem prover for higher-order logic. *TPHOLs 2007 Proceedings – Emerging Trends*, p. 33–48, Kaiserslautern, Germany, 2007.
- [3] C. Berge. *The theory of graphs*. Methuen & Co., 1962.
- [4] F. Brandt et al. The computational complexity of choice sets. *Mathematical Logic Quarterly*, 55(4):444–459, 2009.
- [5] B. Buchberger et al. The Theorema project: A progress report. *Calculus 2000*, p. 98–113, A.K. Peters.
- [6] Common Criteria Recognition Arrangement. Official CC/CEM versions. <http://www.commoncriteriaportal.org/cc/>.
- [7] G. de Clippel, R. Serrano. Bargaining, coalitions and externalities: A comment on Maskin. <http://ssrn.com/abstract=1304712>, 2008.
- [8] C. David, M. Kohlhase, C. Lange, F. Rabe, N. Zhiltsov, V. Zholudev. Publishing Math Lecture Notes as Linked Data. In *The Semantic Web: Research and Applications (Part II of the Proceedings of the 7th Extended Semantic Web Conference ESWC)*, LNCS 6089, p. 370–375. Springer, 2010.
- [9] X. Deng and C. H. Papadimitriou. On the complexity of cooperative solution concepts. *Mathematics of Operations Research*, 19(2):257–266, 1994.

- [10] U. Grandi and U. Endriss. First-order logic formalisation of Arrow's theorem. In *Proceedings of the 2nd International Workshop on Logic, Rationality and Interaction (LORI-2009)*, LNAI 5834, p. 133–146. Springer, 2009.
- [11] T. Hales et al. The Flyspeck Project. <http://code.google.com/p/flyspeck/>.
- [12] J. Harrison. The HOL Light theorem prover. <http://www.cl.cam.ac.uk/~jrh13/hol-light/>.
- [13] E. Hilf, M. Kohlhase, H. Stamerjohanns. Capturing the Content of Physics: Systems, Observables, and Experiments. *Mathematical Knowledge Management*, J. Borwein, W.M. Farmer (eds.), Springer, LNAI 4108, p. 165–178.
- [14] J. S. Jordan. Pillage and property. *Journal of Economic Theory*, 131(1):26–44, May 2004.
- [15] M. Kerber, C. Rowat. Stable sets in three agent pillage games. [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1429326](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1429326), June 2009.
- [16] M. Kerber, Colin Rowat, Wolfgang Windsteiger. Using Theorema in the formalization of theoretical economics, 2011. *Intelligent Computer Mathematics*, J. Davenport, W.M. Farmer, F. Rabe, J. Urban (eds.), Springer, LNAI 6824, p. 58–73.
- [17] A. Kohlhase, M. Kohlhase, C. Lange. Dimensions of Formality: A Case Study for MKM in Software Engineering. *Intelligent Computer Mathematics*, S. Autexier, J. Calmet, D. Delahaye, P. D. F. Ion, L. Rideau, R. Rioboo, A. P. Sexton (eds.), Springer, LNAI 6167, p. 355–369.
- [18] M. Kohlhase. *OMDoc – An Open Markup Format for Mathematical Documents [Version 1.2]*, volume 4180 of LNAI. Springer, 2006.
- [19] M. Kohlhase. Using L<sup>A</sup>T<sub>E</sub>X as a semantic markup format. *Mathematics in Computer Science*, 2:279–304, 2008.
- [20] I. Lakatos. *Proofs and Refutations*. Cambridge University Press, 1976.
- [21] W. F. Lucas. A game with no solution. *Bulletin of the American Mathematical Society*, 74(2):237–239, March 1968.
- [22] E. Maskin. Bargaining, coalitions and externalities. mimeo, June 2003.
- [23] E. Melis, J. Siekmann. Activemath: An intelligent tutoring system for mathematics. In *7th International Conference 'Artificial Intelligence and Soft Computing'*, LNAI 3070, p. 91–101. Springer, 2004.
- [24] J. von Neumann, O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1947.
- [25] T. Nipkow. Social choice theory in HOL. Arrow and Gibbard-Satterthwaite. *Journal of Automated Reasoning*, 43:289–304, 2009.
- [26] T. Nipkow et al. *Isabelle/HOL — A Proof Assistant for Higher-Order Logic*, LNCS 2283, Springer, 2002. updated version online.
- [27] L. C. Paulson. Mechanizing Unity in Isabelle. *ACM Trans. Comput. Logic*, 1:3–32, July 2000.
- [28] G. Sutcliffe. The TPTP Problem Library and Associated Infrastructure. The FOF and CNF Parts, v3.5.0. *Journal of Automated Reasoning*, 43(4):337–362, 2009.
- [29] R. Vestergaard. A constructive approach to sequential Nash equilibria. *Inf. Process. Lett.*, 97:46–51, 2006.
- [30] F. Wiedijk. Formalizing Arrow's theorem. *Sadhana*, 34:193–220, 2009.
- [31] F. Wiedijk. *The Seventeen Provers of the World*, LNAI 3600, Springer, 2006.

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

Alison Pease and Ursula Martin

**Abstract.** Alan Turing proposed to consider the question, “Can machines think?” in his famous article [38]. 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 [34] and Lakatos [22], and consider how their theories stand up in the light of this documented record of informal mathematical collaboration.

## 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 [38] suggested that one way of measuring whether a machine is intelligent is by comparing its behaviour to human behaviour. We explore his proposal in [33]: 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 [17]. 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

---

<sup>1</sup> School of Electronic Engineering and Computer Science, Queen Mary, University of London, Mile End Road London E1 4NS, UK.  
Email {Alison.Pease, Ursula.Martin}@eeecs.qmul.ac.uk

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 [16] he said that despite wide participation initially, the number of contributors settled down to a handful, all of whom he knew personally (Nielsen [31] 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.” [16]. 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 [34] and Lakatos [22], 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 contestants on Monday July 18th. Tao posted the problem at 8pm on July 19th, 2011 Coordinated Universal Time (UTC)<sup>2</sup>, 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.

---

**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  $p \neq p$  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)

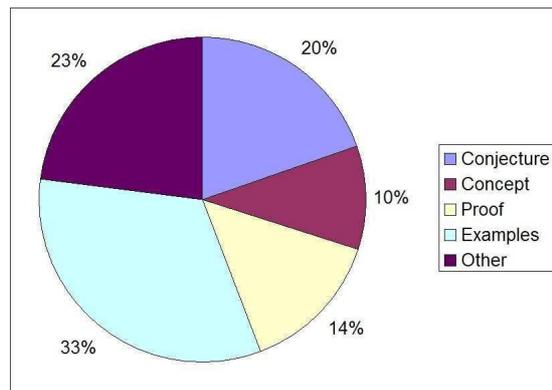
**4a.** The first point and line  $P_0, l_0$  cannot be chosen so that  $P_0$  is on the boundary of the convex hull of  $S$  and  $l_0$  picks out an adjacent point on the convex hull. Maybe the strategy should be to take out the convex hull of  $S$  from consideration; follow it up by induction on 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)

---

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.

<sup>2</sup> All timepoints given hereafter refer to this date.



**Figure 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.

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 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.<sup>3</sup>) 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

<sup>3</sup> This calculation ignores the threads which were simply pingbacks.

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.

---

**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)

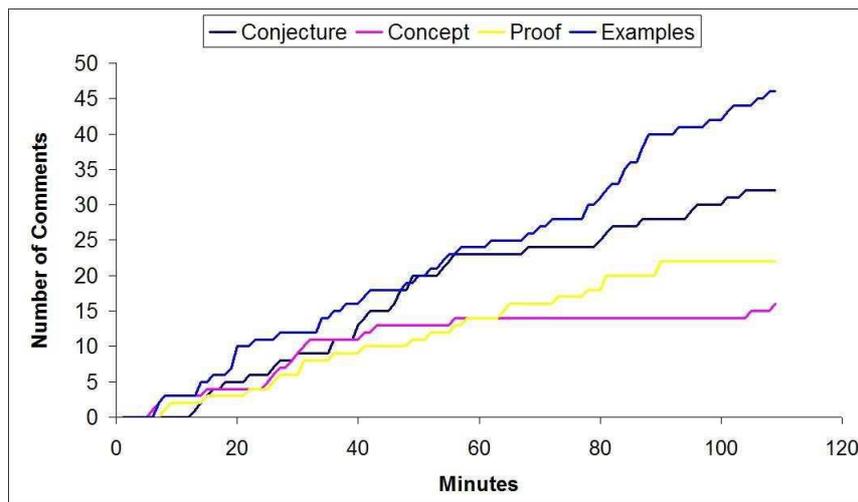
---

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)

---



**Figure 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.

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.

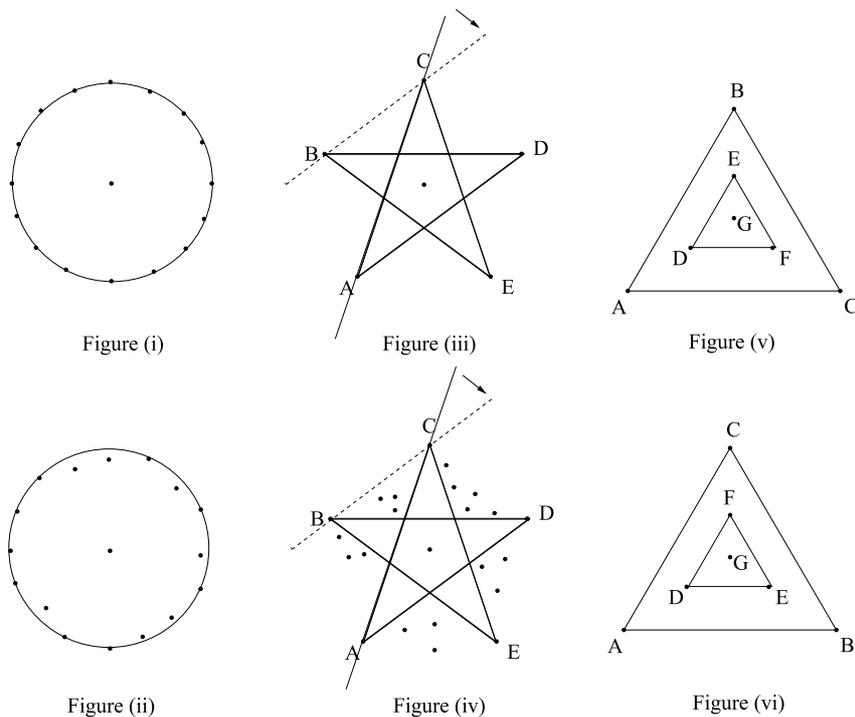
The next comment in the thread, by Joel, is another request for an example: this is related to the first request, with further criteria which

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 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.



**Figure 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.

---

**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 just 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.” [37, 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 discussing 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  $\pi$ . 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 don’t 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 at 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, [8, 12, 25]).

### 10.1 Pólya's problem-solving heuristics

Pólya characterised problem-solving methods, collected in his *How to Solve it* [34], in order to aid the teaching and learning of mathe-

tics. The career of these heuristics has been somewhat chequered, peaking early on with a profusion of problem-solving courses and seminars in which [34] 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 [34]). 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 [36]. 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 [22]. 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 mathematical 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 iden-

tifies 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 [23] and Barton [6] 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.<sup>4</sup>

The main case study in [22] 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 unhelp-

ful – 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.

## 11 A computational blogger

Turing formalised the notion of algorithm [39], and algorithms for finding proofs have been defined and implemented as computer programs in a host of theorem provers, such as Otter [29], Vampire [35] and SPASS [7]. Gowers [18] 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.” [18, 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 [27] 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 [24], which was designed to both construct new concepts and conjecture relationships between them; Colton’s HR system [9, 10], 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 [26], which aims to build a tool which takes in a set of axioms, concept definitions and a logic and

<sup>4</sup> 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.

applies its inference rules to reason from the axioms to theorems; the IsaScheme system by MontanoRivas [30], which employs a scheme-based approach to mathematical theory exploration, and the IsaCosy system by Johansson *et al.* [20] 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 [40] for a history of the proof and controversy), and computational proofs of the Robbins Conjecture by McCune [28] and the Kepler Conjecture [19] 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 [34] has inspired computational accounts, such as emulations of his work on reasoning by analogy [14, 21], a Pólya-inspired explanation of algorithms and data structures [13] and an account of how to select between different problem-solving methods [15];
- Lakatos: The HRL system by one of us [32] is a computational account of Lakatos's theory [22], implemented in an agent architecture where each agent has a copy of the HR system [9]. Agents form and communicate conjectures, and then find and respond to counterexamples, following Lakatos's theory. Colton and Pease's TM system [11] 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 [38] was right when he commented: "We can only see a short distance ahead, but we can see plenty there that needs to be done."

## ACKNOWLEDGEMENTS

We would like to thank Geoff Smith for reading and discussing the paper.

## REFERENCES

- [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](http://michaelnielsen.org/polymath1/index.php?title=Imo_2011_2011).
- [6] B. Barton, *The Language of Mathematics: Telling Mathematical Tales*, Mathematics Education Library, Vol. 46. Springer, 2009.
- [7] Weidenbach C., Dimova D., Fietzke A., Kumar R., Suda M., and Wischniewski P., 'SPASS version 3.5', *22nd International Conference on Automated Deduction*, 140–145, (2009).
- [8] C. Cellucci and D. Gillies, *Mathematical reasoning and heuristics*, Kings College Publications, London, 2005.
- [9] S. Colton, *Automated Theory Formation in Pure Mathematics*, Springer-Verlag, 2002.
- [10] S. Colton and S. Muggleton, 'Mathematical applications of Inductive Logic Programming', *Machine Learning*, **64**, 25–64, (2006).
- [11] 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).
- [12] D. Corfield, 'Assaying Lakatos's philosophy of mathematics', *Studies in History and Philosophy of Science*, **28**(1), 99–121, (1997).
- [13] R. G. Dromey, *How to Solve It by Computer*, Prentice Hall, 1982.
- [14] T. G. Evans, 'A program for the solution of geometric-analogy intelligence test questions', in *Semantic Information Processing*, ed., M. Minsky, 271–353, The MIT Press, Cambridge, Massachusetts, (1968).
- [15] E. Fink, 'How to solve it automatically: Selection among problem-solving methods', in *Proceedings of Artificial Intelligence Planning Systems*. AAAI, (1998).
- [16] T. Gowers. Polymath1 and open collaborative mathematics. <http://gowers.wordpress.com/2009/03/10/polymath1-and-open-collaborative-mathematics/>.
- [17] T. Gowers. Is massively collaborative mathematics possible? <http://gowers.wordpress.com/2009/01/27/is-massively-collaborative-mathematics-possible/>, March 2009.
- [18] W. T. Gowers, 'Rough structure and classification', *GFAFA (Geometric And Functional Analysis)*, **Special volume – GFAFA2000**(1–0), (2000).
- [19] 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).
- [20] M. Johansson, L. Dixon, and A. Bundy, 'Conjecture synthesis for inductive theories', *Journal of Automated Reasoning*, **Forthcoming**, (2010).
- [21] 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).
- [22] I. Lakatos, *Proofs and Refutations*, CUP, Cambridge, UK, 1976.
- [23] G. Lakoff and R. Núñez, *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*, Basic Books, 2000.
- [24] D. B. Lenat, 'Automated theory formation in mathematics', in *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, pp. 833–842, Cambridge, MA, (1977). Morgan Kaufmann.
- [25] P. Mancosu (Ed), *The Philosophy of Mathematical Practice*, Oxford University Press, 2008, USA.
- [26] R. McCasland and A. Bundy, 'MATHsAiD: a mathematical theorem discovery tool', in *Proceedings of SYNASC*, (2006).
- [27] 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).
- [28] W. McCune, 'Solution of the Robbins problem', *Journal of Automated Reasoning*, **19**(3), 263–276, (1997).
- [29] W.W. McCune, 'Otter 3.0 Reference Manual and Guide', Technical Report ANL-94/6, Argonne National Laboratory, Argonne, USA, (1994).

- [30] Omar Montano-Rivas, Roy McCasland, Lucas Dixon, and Alan Bundy, 'Scheme-based theorem discovery and concept invention', *Expert Systems with Applications*, (0), –, (2011).
- [31] M. Nielsen, *Reinventing Discovery: The New Era of Networked Science*, Princeton University Press, USA, 2011.
- [32] A. Pease, *A Computational Model of Lakatos-style Reasoning*, Ph.D. dissertation, School of Informatics, University of Edinburgh, 2007. Online at <http://hdl.handle.net/1842/2113>.
- [33] 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).
- [34] G. Pólya, *How to solve it*, Princeton University Press, 1945.
- [35] A. Riazanov and A. Voronkov, 'The design and implementation of VAMPIRE', *AI Comm*, **15**(2-3), 91–110, (2002).
- [36] A.H. Schoenfeld, 'Pólya, problem solving, and education', *Mathematics Magazine*, **60**(5), pp. 283–291, (1987 (Dec.)).
- [37] G. Smith, 'Observations: The 2011 IMO in the Netherlands', Technical report, (2011).
- [38] A. M. Turing, 'Computing machinery and intelligence', *Mind*, **59**, 433–460, (1950).
- [39] A.M. Turing, 'On computable numbers, with an application to the entscheidungs problem', *Proceedings of the London Mathematical Society*, **42**(2), 23065, (1937).
- [40] R. Wilson, *Four Colors Suffice*, Penguin Books, London, 2002.

# Mathematical Notation and Analogy

Alan Smaill<sup>1</sup>

**Abstract.** The choice of mathematical symbols, such as connectives, while being arbitrary from a logic point of view, is nevertheless important for intuitive understanding, especially in new and unfamiliar theories. The case of Linear Logic, where there are two distinctly different proposals for the choice of connectives, shows the influence of different mathematical analogies at work.

## 1 Analogy via Infomorphism

One way to treat analogy in mathematics is via some notion of structure-preserving morphism from syntactic representations to syntactic representations (or structures to structures, or both). A recent example of this sort of approach is [4].

One way to look for this in action is to look at the choice of symbols (rather than in the axiomatisation directly, for example) when a new mathematical theory or domain is being developed. Here the mathematician is free to introduce not only natural language-like terminology, but also to introduce new mathematical or logical symbols, which are often notational variants on existing symbols, chosen to convey a close relationship with some existing symbol. For example, the integral sign  $\int$  started out as a version of the letter “S”, for sum; the choice of the Greek “ $\Sigma$ ” for summation of series has the same motivation. This suggests that the relation of visual similarity between symbols gives a hint to an analogy between the new domain, for which the symbol is introduced, and some well-established domain, in which the original symbol already plays a role.

Our intention is to use the machinery described above to elucidate this sort of situation; the case of linear logic provides us with some useful information.

An interesting aspect here is that we see a blurring of the line between what is usually thought of as variation in surface presentation (eg different fonts as in  $\Sigma$ ,  $\Sigma$ ; and different symbols as in  $\rightarrow$ ,  $\multimap$ ).

## 2 Case Study: Linear Logic

Girard in 1987 introduced a new logic called Linear Logic that deals with propositions as resources, which can be used up or created. For the standard propositional connectives (and, or, implies), the new system has two new connectives, known as the *additive* and *multiplicative* versions of the connective. Girard’s notation is given in [2].

Subsequently Troelstra’s exposition of Linear Logic [7] uses different notation for some of the symbols, keeping others, and even swapping a couple of Girard’s symbols around. Troelstra explain his motivation for preferring different symbols, and gives his account of Girard’s own motivation.

The choice of propositional connectives is given in table 1.

symbol	Girard	Troelstra
negation	$\perp$	$\sim$
mult conjunction	$\otimes$	$\star$
mult disjunction	$\wp$	$+$
mult one	$\mathbf{1}$	$\mathbf{1}$
mult zero	$\perp$	$\mathbf{0}$
additive conjunction	$\&$	$\sqcap$
additive disjunction	$\oplus$	$\sqcup$
additive one	$\top$	$\top$
additive zero	$\mathbf{0}$	$\perp$
linear implication	$\multimap$	$\multimap$

**Table 1.** Comparison of connectives

We see that in this set of connectives only three are in common ( $\multimap$ ,  $\mathbf{1}$ ,  $\top$ ); Troelstra has different symbols for five, and confusingly the symbols  $\perp$ ,  $\mathbf{0}$  are swapped around between the two presentations.

Because the two presentations are of the same logic, simply using different symbols, in formal terms the axiomatisations are very nearly identical; in terms of signature morphisms they are isomorphic, which is to say that they are notational variants [5]. But, in this unfamiliar setting, having symbols that carry certain expectations from previous experience can have a large influence on how easy it is to understand the system. (Personally, I find Troelstra’s presentation easier to remember, despite having seen and worked with Girard’s first.)

We can consider the two presentations above as involving *design choices* about which symbols to use — to reuse symbols with established, and indeed multiple uses, or invent new symbols which are nevertheless reminiscent of familiar symbols, as in the cases of the implication (the “lolly”:  $\multimap$ ), or even carrying few suggestions (the “par”:  $\wp$ ). [3] discusses such design choices in the context of his so-called semiotic morphisms, a variant on the morphisms mentioned above.

Design is the problem of massaging a source space, a target space and a morphism to achieve suitable quality, subject to constraints. The extent to which structure is actually preserved gives a way to compare the quality of semiotic morphisms ...

Goguen in [6]

The “preservation of structure” alluded to here corresponds to the possibilities for analogical transfer, such as familiar properties associated with a given symbol, or indeed a given similar symbol. The possibility of such analogical transfer is of course one of the distinctive features of analogical reasoning in general. [1].

Consider using the correlation between symbols (the visual similarity) as indicative of a systematic relationship between two sign systems. Thus an analogy where symbols are mapped to similar symbols may support analogical transfer of the appropriate relationships

<sup>1</sup> Edinburgh University, Scotland, email: A.Smaill@ed.ac.uk

more easily than would be the case if dissimilar symbols were involved. This is related to Goguen's criteria for goodness of semiotic morphisms in [6, sec 4], if we allow a weaker notion of preservation of constructors.

## 2.1 Troelstra's choice

Troelstra in [7, pp 21–22] explains his choice of connectives, which emphasise symmetries seen in many logics, based on the de Morgan duality between classical logical disjunction and conjunction, eg the equivalence of  $\neg(P \wedge Q)$  with  $\neg P \vee \neg Q$ ; he thus has these laws holding for the pairs  $\sqcup, \sqcap$  and  $+, \star$ . These depend on the pairs of symbols being associated together in previous situations. The choice of  $\sqcup, \sqcap, \top, \perp$  is motivated by the fact that under a natural choice of order on formulae (when one logically entails the other), these in fact obey the axioms for lattices.

## 2.2 Girard's choice

Girard's choice also invokes previously understood situations, in this case tensor algebra, where  $\otimes$  is used for example for the tensor product of modules;  $\oplus$  is used there for the direct sum, and the distributivity law from that situation holds here:

$$A \otimes (B \oplus C) \text{ is equivalent to } (A \otimes B) \oplus (A \otimes C)$$

From other situations, it is natural here to associate  $\mathbf{1}, \mathbf{0}$  with the identity elements associated with  $\otimes, \oplus$  respectively (eg  $\mathbf{1} \otimes A$  is equivalent to  $A$ ).

## 3 Conclusion

A view on the relationship between different axiomatisations that takes into account similarity of symbols enables us to see the mathematical analogies that are in play more or less easily. There is a cognitive claim here, which would surely be hard to substantiate, suggesting that in situations where a particular analogical transfer would be useful, the notation that emphasises that particular analogy would result in easier understanding, or faster problem-solving. There is an opportunity here to look at the application of Cognitive Modelling techniques to some version of this problem.

## REFERENCES

- [1] D. Gentner, 'Analogy', in *A companion to cognitive science*, eds., W. Bechtel and G. Graham, 107–113, Oxford: Blackwell, (1998).
- [2] J.-Y. Girard, 'Linear Logic', *Theoretical Computer Science*, **50**, 1–102, (1987).
- [3] Joseph Goguen, 'Semiotic morphisms, representations, and blending for user interface design', in *Proceedings, AMAST Workshop on Algebraic Methods in Language Processing*, eds., Fausto Spoto, Giuseppe Scollo, and Anton Nijholt, pp. 1–15. AMAST Press, (2003).
- [4] Joseph Goguen, 'Mathematical models of cognitive space and time', in *Reasoning and Cognition: Proceedings of the Interdisciplinary Conference on Reasoning and Cognition*, eds., D. Andler, Y. Ogawa, M. Okada, and S. Watanabe, pp. 125–148, Tokyo, (2006). Keio University Press. On-line version updated.
- [5] Joseph Goguen and Rod Burstall, 'Institutions: Abstract model theory for specification and programming', *Journal of the Association for Computing Machinery*, **39**(1), 95–146, (January 1992).
- [6] Joseph Goguen and D. Fox Harrell, 'Information visualization and semiotic morphisms', in *Multidisciplinary Approaches to Visual Representations and Interpretations*, ed., Grant Malcolm. Elsevier, (2004).
- [7] A. S. Troelstra, *Lectures on Linear Logic*, CSLI Publications, Stanford CA, 1992.

# Discovery in Mathematics as an Experiential Practice of Privation

Sandra Visokolskis<sup>1</sup>

**ABSTRACT.** The purpose of this paper is to clarify certain issues concerning the notion of creative discovery in mathematics. This is done so by presenting an integrative model of the creative process based on cyclical stages, which give an account of the sudden emergence of mathematical results in a disruptive, rather than in a non automatic nor mechanical, manner.

This leads to the characterization of the discovery in mathematics as an active and experiential process. When such process results in a creative act of a not completely deductive origin, it is obtained due to an instance of privation, i.e. what Aristotle had interpreted in terms of an attribute that a subject could have whether it is yet in its possession or not. Instead, what happens at discovery process occurs through an unexpected twist, namely, that such privation anticipates a breach in the standard reasoning, thus allowing room for creativity. Therefore, in cases of not fully deductive derivation, discovery is dialectic of privation and creation.

## 1 INTRODUCTION

Discovery in mathematics may be interpreted to appear in two different ways at least. It seems that a mathematician reaches an innovative result either by means of a process of sustained reasoning -continuum-, or, due to difficulties that block the normal course of the rational process, in which case the hard work is stopped, and begins breaking away from the task, -intentionally or not-, appearing some time later a solution, with the illusion of being fortuitous -disruptivism-.

This paper attempts to characterize the creative discovery in mathematics, in a disruptive manner, i.e. the second alternative, where the rational break that occurs on the way to solve the problem is generally due to an aporetic situation, i.e., an apparent dead-end is reached, moment which creativity's theorists baptize as "incubation period" [32]. We seek to characterize the incubation as a privation, i.e., a notion that Aristotle described in terms of an attribute that a subject could have, even though it is not in its possession yet, which is not merely an absence but a potentiality that can be changed into being that actual attribute.

Then, incubation, so understood, consists of a period culminating in the creative "illumination" act – as it was called and interpreted by Wallas[32], and that we characterize here as the last instance of an extreme and culminating privation of a process that is reached from a blocked state. Such state acquires

traits which are somewhat similar to that of a traumatic experience, instance that we will call "leap into the void", concept which we will touch upon further below.

But this is an experience that solves the matter as it brings about a solution of the previous conflict. The incubation as privation leads to a kind of experience which must be unraveled in order to arrive to a solution –which is not always achieved, but here we are analyzing the favorable cases -.

It is the purpose of this work to chart a path to disengage the threads of these cognitive mechanisms that are supposedly present throughout different instances of the discovery process applied to the solution of a mathematical problem. This unfolds a more integrative approach focused on mathematical practice as a cognitive experiential activity. In this way, we present a model of the creative discovery process, based on recyclable stages, alternative to the one offered in 1926 by Graham Wallas [32], taking into account its criticisms, in search of a disruptive perspective, explaining phenomena such as incubation, flashes of insight, and other involved not entirely rational cognitive processes.

## 2 THEORY VERSUS PRACTICE, THEOREMS VERSUS PROBLEMS: A PRELIMINARY ELUCIDATION

To sustain our thesis, we begin by assuming the perspective that the discovery process can be characterized by a search for solutions to a problem. When this solution is not as straightforward as it is expected regarding manner, in lesser or greater degree is generally considered to be a unique, unexpected, and outstanding innovative process, i.e. "creative".

However, our proposal will differ in problem solving from the mainstream. The reason being that it requires the consideration of a wide range of interdisciplinary mechanisms: not only logic-inferential structures but also epistemological mechanisms around the setting of prior hypotheses which must be torn down to accommodate innovative proposals. At the same time, there are also added psychological mechanisms intermingled with the mental blockages experienced after a long and unsuccessful work routine. In addition, also mechanisms appear inherent to the ways in which the mathematical creator manages to get rid of fixations to ideas previously conceived and taken as the working norm. Here, the idea of privation prevails, as we shall see below. All these elements, and some others which will be addressed later, are not usually considered in a standard approach to problem solving, at least not in a disruptive style, as it is proposed here.

<sup>1</sup> Dept. of Philosophy, Nat. Univ. of Córdoba, Argentina, email: [sandravisov@gmail.com](mailto:sandravisov@gmail.com)

Before we move on, we need to clarify that in the history of mathematics, as early as the 4th century AD, Pappus, in the framework of the ancient Greek tradition discussed two ways of conceiving the discovery process -process which he also called of "analysis"- regressive and opposed to the progressive "synthesis", considering the latter as a systematization of the first: problems and theorems<sup>2</sup>.

Even though, viewing mathematics only as a matter consisting of theorems and problems tends to be a very simplified perspective on the matter at hand, it helps us to understand the priority in the theorems which reached the philosophical perspective of mathematics of ancient Greece. This is a perspective which emphasizes the issue of proof -and thereby the deductive method- as the core of the discussion on the progress of mathematics as theoretical science. Mathematics, characterized as the science of theorem proving, is then the hegemonic discipline of the search for reasoning.

We should wait for authors such as George Polya (in 1945) [17] and Imre Lakatos (in 1976) [14] to re surface the issue of problems in mathematics and heuristic methods of discovery. As a consequence, this will generate a new tradition around problem solving in mathematics. Years later, in the path traced by the authors cited, we are closer to the current emergence of mathematical practice<sup>3</sup>.

Here, the distinction "theory versus practice" in mathematics is not necessarily reduced or settled in the complicated demarcation between theorems and problems. A philosophical approach underwritten by the latter as the core of everyday mathematical work, which puts emphasis on the specific activity of the mathematician and not so much on philosophical speculation sometimes detached of actual practice.

Both concepts are two sides of the same coin. This is a clear display of different, yet complimentary, styles of carrying on: one mostly synthetic and the other mainly analytic, respectively. However, the cognitive mechanisms that apply in one or other case, as far as they involve analytical or synthetic creative processes, may differ in content but not in form. Consequently, this assertion stresses that while synthesis consists on the creation of deductive strategies, in the case of analysis, analyses itself is an unnecessary ban and is usually ruled out.

The description presented above can be summarized by saying that there is creativity in both what usually is called the context of discovery, as well as on the systematization of deductive processes relevant to the context of justification. There could be as much creativity in finding innovative results as in trying to provide new deductive justifications.

What is relevant is that we are talking about simpliative knowledge in both cases: in the first one,

because there new conclusions, previously unforeseen, arise that require to be warranted the results in a post hoc way; in the second, because demonstrations never before applied come out. Nevertheless, it is clear that mathematical research is and should be presented in theorematic terms.

In this regard, we have traces in the Alexandrine Greece of writings addressed to Eratosthenes where Archimedes<sup>4</sup> addresses Eratosthenes expressing a clear, latent, demarcation between demonstrative methods -as it is the case of the method of exhaustion<sup>5</sup>- and methods created to find results - as it is in the case of his proposal for the mechanic method-, although the latter were not entitled as above. However, both methods show important creative aspects. Archimedes says:

I thought fit to write out for you and explain in detail in the same book the peculiarity of a certain method, by which it will be possible for you to get a start to enable you to investigate some of the problems in mathematics by means of mechanics. This procedure is, I am persuaded, no less useful even for the proof of the theorems themselves; for certain things first become clear to me by a mechanical method, although they had to be demonstrated by geometry afterwards because their investigation by the said method did not furnish an actual demonstration. (Heath, 2002:12-13)

In what follows, we will describe a model of creativity applied in both contexts, emphasizing the development of an explanation that gives an account of the creative discoveries. We are aware that such discoveries are not necessarily all the types of discoveries that happen but in this paper we will focus on the ones where some not entirely rational cognitive mechanisms emerge. We shall keep in mind that creativity, according to our disruptive perspective and, therefore, as an experiential practice of privation, requires a breach in the standard reasoning.

### 3 A PREVIOUS ISSUE ABOUT REFORMULATION OF PROBLEMS

As we stated above in 1, the disruptive point of view is based on an incubation interpreted as privation, which converges to a extreme state of rational break. But what causes such break? The continued and sustained search for a solution to the problem in ways that ultimately are unsuccessful in some key aspect. Then, what happens before this obstruction? Two general responses are possible: a refusal -which here we omit as it is not relevant- and a positive response that offers a radical reconsideration of the problem,

<sup>2</sup> See Book VII of Pappus's *Mathematicae Collectiones*, where he writes about the "Treasure of Analysis", and discusses analysis and synthesis in geometry. For a discussion of this topic, see [26], [12] on Pappus, and [10] and the introductory notes to Euclid's *Elements* from [24], on problems vs. theorems and analysis vs. synthesis.

<sup>3</sup> It's worth noting the comment Eduard Glas [6] makes in connection with the use Lakatos makes of the discoveries in mathematics in a theory-centered sense, concerning about the "logic of discovery", therefore, not becoming a true propagator of a philosophy of mathematical practice.

<sup>4</sup> See [3], [4], [8], [11], [23], and [25].

<sup>5</sup> Although the term "exhaustion" was first introduced by Grégoire de Saint-Vincent in 1647, Archimedes ascribed this method to Eudoxus, who proposed a rigorous inferential strategy, as given in Euclid's *Elements* XII, avoiding what we today call the direct passage to the limit, which was not clear at that time. See [4].

which will open the possibility of a solution. In this section, we will deal with this issue.

If discovery in mathematics is a form of problem solving, as we stipulate in section 2, apparently, every discovery process might occur only in relation to a set of prefixed initial data and goals that give rise to the problem's constraints.

This is a central point of discussion because we precisely describe the final creative act or insight as a moment in the whole process of discovery in which the set of assumptions is discarded in order to sustain other hypotheses. This allows an innovative reformulation of the problem, one that now does provide a solution to it.

Therefore, the initial problem can and should be reinterpreted in view of the need to reflect the researcher's reconstructed intentions according to the new demands of the task. In this regard, in relation to the translation of a geometric problem into algebraic equations, George Polya (1962) understands the problem solving process, even if it's not necessarily creative, as a dynamic of constant reworking:

As our work progresses, our conception of the problem continually changes: more and more auxiliary lines appear in the figure, more and more auxiliary unknowns appear in our equations, more and more materials are mobilized and introduced into the structure of the problem till eventually the figure is saturated, we have just as many equations as unknowns, and originally present and successively mobilized materials are merged into an organic whole. (Polya, 1962: 57)

Consequently, the original problem, as it was established in the past, will be rethought in the present from the provision of new tools. However, the idea of past memories is not appropriate if we understand the recovery processes as recovery information stored in old slots of memory. Against this, we adhere to the thesis of the intentional actions that Ian Hacking [8] outlined in the following terms:

When we remember what we did, or what other people did, we may also rethink, redescribe, and refeel the past. These redescriptions may not have been true in the past, that is, not truths about intentional actions that made sense when the actions were performed. That is why I say that the past is revised retroactively. What was done itself is modified. The past becomes filled with intentional actions that, *in a certain sense, were not there when they were performed*<sup>6</sup>. (Hacking, 1995: 249-250)

This task of reconceptualization, nevertheless, does not tend, at first, to be easy. The reason being that it requires that the mathematician is predisposed to overcome his/her resistance to reinterpreting the original hypotheses. It also calls for a broad mind and implicit acceptance

that certain aspects of mathematics are not eternal, irremovable and monotonically conservable. This implies "not to think of the past as fixed, final and determined" (Hacking, 1995:234).

It is frequent to hear obsolete characterizations of mathematics as a cumulative science. They are based on enduring and permanent truths that only can produce more truths with the deductive machine. Although it is clear that we start with hypothesis, temporarily acceptable conjectures, until we find not sufficient evidence for its rejection. Precisely what we are talking about now is of the possibility of carrying out a *retroactive redescription* of the problem in question. As Hacking proclaims,

We rewrite the past, not because we find out more about it, but because we present actions under new descriptions<sup>7</sup>. (Hacking, 1995: 243).

Hence, what appears to of relevance for the process of discovery is to be able to transform the previous background information so as to gain at least one of multiple new perspectives on the task at hand. The type of transformation we are talking about provides an instance of *privation*.

Indeed, to achieve a creative breakthrough in solving the problem, necessarily, we must go backwards. Metaphorically speaking, we could compare it to the situation of an athlete who takes a step backwards to thrust forward more energetically in a race, or equivalently, in the art of the archery, which requires tightening the rope in order to have enough negative force and throw the arrow forward.

Here, privation is a requirement for creation, in the same way that when we affirm that "less is more", i.e., we have to create a vacuum of some kind to clarify before a possible contribution to the solution of the problem. Privation, *stéresis*, a term that Aristotle systematized as one of three basic principles of nature (matter and form being the other two), which is the absence of a given form in something capable of possessing it. It is not a cause because privation is insufficient as an element for the reception of the form, and this implies, within Aristotle's framework, that it doesn't have explanatory power, but rather that it is a necessary condition for change, i.e. the terminus a quo of change. We apply here the sense of the term Aristotle uses in *Categories* to indicate some quality, habit, disposition or faculty that should be in some grade in a person, but is absent. Aristotle says:

A privation cannot be affirmed of a subject unless its opposite habit could be naturally

---

<sup>7</sup> From a philosophical point of view, note that what Hacking is talking about is inserted in the preference for an Interventionism that we share at the expense of a Platonism in mathematics, where the latter describes an immutable and eternal mathematical world, where everything, once accepted, exists forever. Instead, the Interventionism of Hacking allows flexibility in the acceptance of results, which play a conjectural, hypothetical and transitional role, and can, eventually, be removed.

---

<sup>6</sup> The italics are ours.

possessed by it. (Aristotle, 1995, 10, 12a, 27-34)

Here, privation is the absence of presence, with no positive or independent existence. Applied analogically to our situation, we are assuming that in order to find a creative solution, a mathematician must potentially have the ideas in a not yet present state, as a privation, momentarily lacking a solution to it. Accordingly, the momentary absence of a solution could be characterized as one in which a solution has not been found because there is no room for its existence yet. Similarly, in mathematics, when we refer to variables and imagine them as containers deprived of a constant that takes their place, i.e. that is how Aristotle outlined the notion of privative ideas.

We will see in the next section how this Aristotelian concept is used here with the purpose of explaining the less rational cognitive aspects of the creative process. We anticipate now that the idea of a break after a routine of hard work that reaches a saturation limit is often characterized as a stage in the creative process that receives the name of "incubation", a term coined by Wallas [32] in 1926, which we will characterize in terms of privation.

Graham Wallas is credited as being the first to formulate a four-stage process of creativity:

- (a) Preparation: where the problem is investigated thoroughly by observing, listening, asking, reading, collecting, comparing, contrasting and analyzing previous materials, knowledge and information seemingly relevant to the task in question.
- (b) Incubation: time during which the task is left aside, not being worked on at a conscious level, but at a preconscious or unconscious one.
- (c) Illumination: a stage where suddenly appears into consciousness a solution in a flash of insight, an "aha" experience, the "eureka" effect.
- (d) Verification: in which the idea must be tested checking its efficacy to further applications.

Despite being a very successful model, especially quoted as a milestone in the notion of incubation, term coined by him, and of illuminative insight, its linearity has been largely discredited. Additionally to this last issue, our concern with this model has to do with the vague sense that Wallas attached to the idea of suddenness (abruptness) of insight that will be exposed below.

According to our interpretation, incubation emerges in the metaphorical description where the rope starts to stretch with the arrow going in direction of departure, or in the case of the athlete, just when the foot is going backwards.

The seeds of problem solving paradoxically are in what a conscious mind fears the most: the loss his/her mind, being out of control, and along with that, to accept vulnerability, and precisely in this aspect is what incubation and privation are all about, i.e. the capacity to surrender to one's previous formulation of the problem, an experience that must be undone. It involves less frontal activity, which facilitates disinhibition of the sub-cortical regions and can evoke feelings of dissociation, fantasies even. As a consequence, this allows the subject to be dynamically able to engage several brain regions and

have fluent associations, rapidly shifting between different neural circuits, provoking broad attentional focus.

Note that the acceptance of vulnerability is not without fault on the subject. According to Psychoanalytic theories that reinforce the idea that a subject is considered "normal" as long as it retains his sanity, i.e. exercising reason permanently (except perhaps when he sleeps) and assuming the attribution of his acts. Regarding this matter, the subject assumes that it should always conform to the rule of reason which consists in "responding to others". This interpretation of an individual responsibility, with social punishment for those who do not comply with that rule, seems corrupted when the subject becomes absorbed in thought, lapsed experiences of some nature, as it occurs when entering incubation.

It should be noted that the creative process has the characteristics of a paradoxical situation described as the dialogue between opposite ends: not just about privation versus creation but also a balance between emotion and reason. Given that during the incubation phase the tip of the balance favors the intensification of emotions at play, at the expense of a slowdown in intellectual capacity. This leads to a stand-by state over the issue in question. As we shall see below, these emotions reach the peak of crisis during stage 3 of our interpretation of the creative process, state of blocked work, causing a near-trauma affective experience, which gives way to incubation (stage 4 of our interpretation).

The kind of trauma we are thinking about here is a disturbing experience; a state which accompanies stressful situations influenced by threats to his/her research, and also, a lack of capacity to address that threat which generates such inability to respond. It involves also negative emotions, diminution of concentration and desmotivation, including overwhelming anxiety, occasionally, to the point of being burned-out.

This assertion may be shocking, given the pejorative connotations commonly given to the term "trauma", and the idea of something derogatory and abnormal, not usually associated with people in their right mind. It would seem that we are describing the discoverers with a touch of madness, but, as we shall see below, madness is not necessarily a consequence of assuming our approach.

In fact, although, at the beginning, the term was associated in with bodily injury, later on, it acquired a psychological connotation, as Hacking says:

Trauma took the leap from body to mind just over a century ago, exactly when multiple personality emerged in France, and during the time when the sciences of memory were coming into being. (Hacking, 1995: 183).

This traumatic instance is generated by a break in the continuous mental productions; situation which some authors think is unquestionable<sup>8</sup>. As a result, these "breaks allow redirection of attention from irrelevant aspects of the problem to the relevant ones" (Neçka, 2011: 669), admitting the possibility of

---

<sup>8</sup> See [15].

reformulating the problem to finally come up with a solution for it, clearly not based entirely on the same problem but another highly modified one on the basis of mental openness. From the Gestalt Theory, Stellan Ohlsson says in this respect that:

Restructuring means that a new problem arises in one's mind, with new initial data, new goals, and new constraints. When such a change is completed, the transformed mental representation of the problem works as a fresh source of retrieval cues, thanks to which some relevant data, already stored in the long-term memory, can be used in a productive way. (Nečka, 2011:669-670)

In short, the reformulation of the problem is vital in the pursuit of its resolution when such problem is blocked. The key is to avoid the paralysis which the subject usually faces. In order to achieve a result, he/she will have to surrender any kind of absolutist claims of resolving the original problem without changing the starting assumptions. Which, paradoxically, does not mean to continue to do what was done previously, capriciously maintaining all the initial conditions of the problem, but to take one step back in the course of action taken before, which in turn will allow incubation. Therefore, this is a partial "let it be" reformulation (our stage 8), which culminates in a phase of systematization of the solution (stage 9), and these two stages conform the process of synthesis.

#### **4 THE PROPOSAL: A COGNITIVE DISRUPTIVE MODEL OF CREATIVE DISCOVERY**

On the way towards an explanation of the process by which incubation allows access to the creative insight, within our disruptive model of creative discovery framework, we briefly summarize all stages and their meaning. In doing so, we are aware of the fact that not all discoveries in mathematics can be called "creative", but wherever the discovery reaches certain level, as we describe above, creativity is at work:

##### **I. Analysis:**

Stage 1: Original formulation of the problem. A clarification of the initial data, the goals and the constraints.

Stage 2: Systematic hard work. Persistent search for solution. The problem solver explores all the paths at hand, exhausting all possibilities. As Feldman affirms: "It seemed as if my mind were designed to do nothing else" (Feldman, 1988:187). Or, as Andrew Wiles expresses himself referring to Fermat's problem: "I carried this problem around in my head basically the whole time,...without distraction,...[like] one particular thing buzzing in [my] mind,...I was thinking of it all the time." (Wiles, 2012). Polya also contributes to this idea: "'Genius is patience'. 'Genius is one per cent inspiration and ninety-nine per cent perspiration.' One of these sayings is attributed to Buffon, the other to Edison, and both convey the same message: a good

problem solver must be obstinate, he must stick to his problem, he must not give up." (Polya, 1962: 92). Progressive work: to extract consequences from initial assumptions, working forward, i.e. taking as a starting point the data working its way up to derive to a solution. Regressive work: to search for the reduction of the problem (*apagogé*), working backwards, towards the data, maybe arriving at an auxiliary problem which we can solve.

Stage 3: Blocked work. Incessant work slowed (and even blocked) by some kind of obstruction not determinate it yet. The scientist is struggling with the problem, feeling stuck, with fixed ideas impossible to mobilize, situation which results in a lack of progress, leading up to an affective impasse. The problem drives the researcher to an aporetic situation in which the knowledge previously acquired brings along unexpected complications<sup>9</sup>.

Stage 4: Incubation. An impasse to "put the problem out of mind, while paradoxically remaining sensitive to stimuli and ideas that might be related to the unsolved problem." (Smith, 2011:657)<sup>10</sup>.

Stage 5: Leap into the void. The incubation's final effect: an extreme case taking impact of implementing the incubation to the limit, reaching a state of absence of reason. An experience which increases the feeling that there is no way to escape from the rapid decline of the rational activity; threat of imminent danger upon completion of a dark cycle, accompanied by an overwhelming feeling of being dominated, feeling powerless against external forces beyond ourselves. extreme dissociative state (not pathological), with severe emotional exacerbation and decay of the attentional focus, which in turn, results in an experience of minimum, or even non-existent, awareness, carrying a sensation of vertigo. Metaphorically, it evokes the dizzying fall by the edge of reason, a penetrating but fleeting experiential void in consciousness, as if in total darkness, reaching an instantaneous recovery of the intellectual forces, accompanied by an enlightening flash -the sudden emergence of an omni-embracing conceptual clarity- i.e. illumination, the next stage.

Stage 6: Illumination. Sense of cognitive certainty about the problem and a solution of it. Experience of accessibility to the full knowledge of the problem, in a direct, immediate and unlimited way. However, this doesn't mean that he/she has formal detailed information on technical aspects of the problem, but rather a kind of revelation about it, where the need of reasoning is attenuated in favour of a type of experiential understanding. In this regard, Polya says: "The solution of a problem may occur to us quite abruptly. After brooding over a problem for a long time without apparent progress, we suddenly conceive a bright idea, we see daylight, we have a flash of inspiration. It is like going into an unfamiliar hotel room late at night without knowing even where to switch on the light. You stumble around in a dark room, perceive confused black masses, feel one or the other piece of furniture as you are groping for the switch. Then, having found it, you

<sup>9</sup> See [28].

<sup>10</sup> For an enlarged view of this topic, confront [26] and [30].

turn on the light and everything becomes clear. The confused masses become distinct, take familiar shapes, and appear well arranged, well adapted to their obvious purpose. Such may be the experience of solving a problem; a sudden clarification that brings light, order, connection, and purpose to details which before appeared obscure, confused, scattered, and elusive." (Polya, 1962: 54).

"After the coming of the idea we *see more* -more meaning, more purpose, and more relations. The coming of the idea is similar to switching on the light in a dark room." (Polya, 1962: 59).

Stage 7: Metaphoric Insight. The searcher evokes a metaphor-like statement, which unveils the solution. Patrick H. Byrne refers to this moment stating that: "there is an act of recognition that not only understands the diagram in a new way, but simultaneously recognizes the applicability of the theorem to the diagram; recognizes a 'universal shown through the clear being of the particular' ([1], [2], 71a8)." (Byrne, 1997:116). Polya realizes this fact as well: "A suddenly arising idea, a spectacular new element amid dramatic rearrangement, has an impressive air of importance and carries strong conviction. This conviction is expressed by such exclamations as 'Now I have it!' 'I have got it!' 'That's the trick'." (Polya, 1962: 60).

## II. Synthesis:

Stage 8: Reformulation of the problem.

Stage 9: Systematization of the solution<sup>11</sup>.

There may be recursions before a creative work is produced. This sequence of orderly steps could be entirely, or partly, recycled more than once, as well as be interrupted and reinitiated taking some extra time in any of its phases<sup>12</sup>.

## 5 STEPS TOWARDS CREATIVE INSIGHT

On the process to the reformulation of a problem, where the foundations of our beliefs are shaken, we cross through aporetic situations (stage 3 of blocked work), which usually show the presence of repressed material, that must have been preconscious and needs to be mobilized, recapturing a trend of thought until we reach awareness.

While any kind of routine work is easily achieved, any form of creative activity is associated with extreme difficulties and accompanied by a number of severe physical and mental symptoms. (Kris, 1964: 298)

Any released repression –and, therefore, any possibility of access to the solution of the problem– depends on the strength of the defenses used for the purposes of coping with the prevailing intensity of conflicting trauma. Later on, we arrive to freedom from conflict.

We need to remove obstacles to allow the emergence of new ideas: in these instances, conscious voluntary effort or concentration of attention does not always succeed in recapturing elusory thought. Therefore, suspension of ego effort in an incubation period is at stake: "the experience of clarification that occurs when after intense concentration the solution to an insoluble problem suddenly presents itself following a period of rest." (Kris, 1964: 313) Hopefully, after this stage, the investigator has preconsciously established a unity of context with a reformulation of the problem in hand. In this way, establishing links in a chain of thought which connect firmly together the new hypotheses with the solution in a logic concatenation. This connection fills the gap produced between the old hypotheses and an expected solution, fitting everything into a new frame of reference:

In solving problems, the feeling of fitting propositions together satisfies the requirement of the synthetic function...the psychic concatenation, or the establishing of the unity of context, is due to the synthetic function of the ego. (Kris, 1964: 306-318)

Finally, once the solution is obtained, along with a sense of satisfaction with the achieved task, come a stage of passivity due to a relaxation from tension, as Kris pointed out:

The feeling of triumph and release from tension remind the individual of the phase in his development in which passivity was a precondition of total gratification. (Kris, 1964: 317)

This feeling of passivity is reflected in the idea, often sustained by creators that the solution to the problem, as long as it is original and characteristic of the individual, does not belong to them entirely. This feeling, a mixture of shame and alienation from such solution, has its roots on the lack of understanding of the origin of the idea, which generates a sensation that the result of their creation came to them as a work of fate, or chance. At this instance, the creator feels that the solution appeared as a result of divine providence, as if it were a magic or risky solution, or inspired of by muses, or even as if his/her unconscious was at times more powerful than his conscience, bringing about "unsuspected" solutions to his/her mind.

So, because of this, incubation and its final moment, the leap into the void as it was previously described here, sometimes, causes amazement due to the momentary loss of rational time upon the problem. Such loss, at first, is of an unknown length of time. Also, on that incubatory period, one puts parentheses on its responsibility to the work, thus giving rein to one's imagination and releasing the hypotheses. As says Kris thereon:

The absolution from guilt for fantasy is complete if the fantasy one follows is not one's own. Opportunity for discharge or catharsis is guiltlessly borrowed...One of the main effects of inspiration in primitive

<sup>11</sup> Stages 8 and 9 are developed further in section 2 of this paper.

<sup>12</sup> The need for periods demanding time to process the problem, not just at the incubation period, is cited in several researches. See [19]

society is that of relieving the feeling of anxiety and thus of appeasing the guilty feelings connected with creation. (Kris, 1964: 300-315)<sup>13</sup>

This is why, after arriving to a solution, there is a quest of the mathematician for response by others, i.e. the acknowledgment by response is essential to confirm their own belief in their work, "and to restore the very balance which the creative process may have disturbed. Response of others alleviates the [mathematician's] guilt". (Kris, 1964: 60)

## 6 CONCLUSIONS: THE DISRUPTIVE CHARACTER OF CREATIVE DISCOVERY

The inflection point in our proposal consists in the description of the leap into the void prior to enlightenment, and hence, to the elucidation of the problem, when the intellectual decline that began during incubation reaches its climax. Many specialists agree on the premise that there is a remarkable change, while others argue that it is only apparent. Despite this fact, those who claim that this change is only fictitious agree that it is possible to trace a sequence of steps that would unify the discovery process entirely a posteriori.

Moreover, this position is what we call "continuism"<sup>14</sup>, opposite to ours, namely "disruptivism".

Kris affirms:

The work of the mind in research and discovery does not consist only in a continuous application to the quest for a solution. A part of the work is done in preconscious elaboration, the result of which comes into consciousness in *sudden advances*<sup>15</sup>. (Kris, 1964: 296)

The problem arises when we have to think of how to interpret these "sudden advances". From the perspective of continuism, these sudden advances are actually part of a long chain of previously existing associations but that require the creator to retrieve the steps of the discovery. Hence, the task of the researcher consists on filling the blanks caused by the lack of knowledge of hidden results, which should be unraveled. This implies that the solution may seem to suddenly emerge as if it was disconnected from the problem, but actually, in this view, it is mere appearance, an illusion. Thus, the creator must help in the completion of the process of tracing back the steps, the registered instances which were not previously modified. The creative act is the last step on an incremental process of accumulation of knowledge, only in appearance unfit in kind to the rest of the process.

On the other hand, disruptivism claims that these

sudden advances may only emerge into consciousness after experiencing a peculiar and specific kind of privation: a leap into the void. This plays the role of liberating the researcher from the obstacles to arrive to the solution. These are the very obstacles which were present for him/her during the formulation of the original problem and in the set of hypotheses assumed as certain for its resolution.

Thus, according to disruptivism, there is not a genuine continuity between previous assumptions of the problem in question and its solution. These are misleading assumptions that must not only be modified but abandoned entirely. They must be left on hold to give rise to new versions of them that are connected with the obtained solution but not necessarily with the previous hypothesis of origin.

Therefore, not only is there no appearance of disconnection, but such connection does not exist at all if the result is truly creative. Mind games are what lead us to the desire of filling blank spaces, once the solution is found, in an effort of intellectual recapitulation.

## REFERENCES

- [1] Aristotle, 'Categories', In *The Complete Works of Aristotle*, 2 vols, J. Barnes, Transl. J. L. Ackrill, Princeton University Press, Princeton, (1995).
- [2] J. L. Ackrill, *Aristotle, Categories and De Interpretatione*, Clarendon Press, Oxford, (1963).
- [3] J. Babini (Ed.), *Arquímedes El "Método"*, Editorial Universitaria de Buenos Aires, Buenos Aires, (1966).
- [4] C. B. Boyer, *The History of the Calculus and its Conceptual Development*, Dover Publications, New York, (1949).
- [5] P. H. Byrne, *Analysis and Science in Aristotle*, State University of New York Press, Albany, (1997).
- [6] D. H. Feldman, 'Creativity: Dreams, Insights and Transformations', In *The Nature of Creativity, Contemporary Psychological Perspectives*, R. J. Sternberg (Ed.), Cambridge, Cambridge University Press, (1988).
- [7] E. Glas, 'On the Role of Thought-Experiments in Mathematical Discovery', In: *Models of Discovery and Creativity*, J. Meheus and T. Nickles (Eds.), Springer-Verlag, New York, pp. 57-64, (2009).
- [8] P. M. González Urbaneja y J. Vaqué Jordi (Eds.), *Arquímedes El Método Relativo a Teoremas Mecánicos*, Publicaciones de la Universidad Autónoma de Barcelona, Edicions de la Universitat Politècnica de Catalunya, Barcelona, (1993).
- [9] I. Hacking, *Rewriting the Soul. Multiple Personality and the Sciences of Memory*, Princeton, Princeton University Press, (1995).
- [10] T. L. Heath (Ed.), *A History of Greek Mathematics. Volume I: From Thales to Euclid*, Dover Publications Inc., New York, (1981).
- [11] T. L. Heath (Ed.), *The Works of Archimedes*, Dover Publications Inc., New York, (2002).
- [12] A. Jones (Ed.), *Pappus of Alexandria, Book 7 of the Collection*, part 1 & part 2, Springer-Verlag, New York, (1986).
- [13] E. Kris, *Psychoanalytic Explorations in Art*, Schocken Books, New York, (1964).
- [14] I. Lakatos, *Proof and Refutations. The Logic of Mathematical Discovery*, J. Worrall and E. Zahar (Eds.), Cambridge University Press, Cambridge, (1976).

<sup>13</sup> See on this topic [28] and [30].

<sup>14</sup> Related scholars to this approach of creativity are, e.g., David Perkins, Robert Weisberg and Douglas Hofstadter, among others.

<sup>15</sup> The italics are ours.

- [15] J.Meheus and T.Nickles (Eds.),*Models of Discovery and Creativity*, Springer-Verlag, New York, (2009).
- [16] E. Neçka, 'Insight', In *Encyclopedia of Creativity*, Second Edition, Two Volume Set, M. A. Runco and S. R. Pritzer (Eds.), Elsevier Inc., San Diego, 667-672 (2011).
- [17] G.Polya, *How to Solve it*, Princeton University Press, Princeton, (1945).
- [18] G. Polya. *Mathematical Discovery. On Understanding, Learning and Teaching Problem Solving*, Combined Edition, John Wiley & Sons, Inc., New York, (1962).
- [19] M. A. Runco and N. Cayirdag, 'Time', In *Encyclopedia of Creativity*, Second Edition, Two Volume Set, M. A. Runco and S. R. Pritzer (Eds.), Elsevier Inc., San Diego, 485-488 (2011).
- [20] M. A. Runco and S. R. Pritzer (Eds.), *Encyclopedia of Creativity*, Second Edition, Two Volume Set, Elsevier Inc., San Diego, (2011).
- [21] S. M. Smith, 'Incubation', In *Encyclopedia of Creativity*, Second Edition, Two Volume Set, M. A. Runco and S. R. Pritzer (Eds.), Elsevier Inc., San Diego, 653-657 (2011).
- [22] R. J. Sternberg (Ed.), *The Nature of Creativity, Contemporary Psychological Perspectives*, Cambridge, Cambridge University Press, (1988).
- [23] L. Vega Reñón (Ed.), *Arquímedes: El Método*, Alianza Editorial, Madrid, (1986).
- [24] L. Vega Reñón (Ed.), *Euclides. Elementos. Libros I-IV*, Alianza Editorial, Madrid, (1991).
- [25] P. Ver Eecke, *Les Oeuvres Complètes D'Archimède avec une Introduction et des Notes*, Desclée, De Brouwer & Cie, Paris, (1921).
- [26] P. Ver Eecke, *Pappus d'Alexandrie: La Collection Mathématiques avec une Introduction et des Notes*, 2 volumes, Fondation Universitaire de Belgique, Albert Blanchard , Paris, (1933).
- [27] S. A. Visokolskis, *Modelos de Creatividad Centrados en el Ocio: Una Caracterización*, XX Jornadas de Epistemología e Historia de la Ciencia, Córdoba, (2009a).
- [28] S. A. Visokolskis, *La Poética de Aristóteles. Creatividad y Catarsis*, Universidad Nacional de Córdoba, (2009b).
- [29] S. A. Visokolskis, *Trauma y Creatividad desde el Psicoanálisis Freudiano*, Universidad Nacional de Córdoba, (2009c).
- [30] S. A. Visokolskis, *Hacia un Modelo de la Creatividad basado en la Filosofía Trágica Griega: El Caso Aristotélico*, Universidad Nacional de Córdoba, (2010).
- [31] S. A. Visokolskis, *Creatividad, Atención y Disociación*, Workshop de Psicología, Universidad Nacional de Córdoba, (2011).
- [32] G. Wallas, *The Art of Thought*, Harcourt, New York, (1926).
- [33] A. Wiles, *Interview: Andrew Wiles on Solving Fermat*, Available at: <http://www.pbs.org/wgbh/nova/proof/wiles.html>, (2012).