# D4.1
# Examples of social creativity



| | |
|---|---|
| Authors | Joseph Corneli, Alison Pease, Danae Stefanou |
| Reviewers | Alan Smaill, Ewen Maclean |





| | |
|---|---|
| Grant agreement no. | 611553 |
| Project acronym | COINVENT - Concept Invention Theory |
| Date | September 30, 2014 |
| Distribution | PU/RE/CO |

**Disclaimer**



The project COINVENT acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open Grant number 611553.

# Abstract

This deliverable presents examples of mathematical dialogue and examples of frameworks for thinking about mathematics. The examples are drawn from real-time and text-based online problem solving interactions, and are supplemented by brief examples of single-author works in order to provide a comparison case. Theoretical approaches are divided between dialectics, logic, and pragmatics. The methods are qualitative, and oriented towards the design of future systems and experiments. We include a comparison with social creativity in music to illustrate both the range of applicability of our approach and some of its limitations.

Keyword list: **Inference Anchoring Theory**, **Lakatos**, **blending**, **speech acts**, **mathematics**, **MiniPolymath**, **improvisation**.

## Changes

| Version | Date | Author | Changes |
|---|---|---|---|
| 0.1 | 10.09.14 | Joseph Corneli | Creation |
| 0.2 | 30.09.14 | Joseph Corneli | Respond to reviewer comments |
| 0.3 | 30.09.14 | Joseph Corneli | Final changes from Danae |

## Executive Summary

In this paper we are concerned here with both dialogue and logic, and with connections between the two. We focus for the most part on the mathematics domain. We include a short discussion of related issues in music that helps to illustrate both the range of applicability of our approach and some of its limitations.

The primary ingredients in our formal theory of social creativity are Barwise and Seligman's Information Flow, Goguen's Unified Concept Theory, and Lakatos's Proofs and Refutations. In the broader features of the model we are also inspired by the work of Katarzyna Budzynska and Chris Reed on Inference Anchoring Theory (IAT), which develops connections between a dialogical context and logical argument by way of illocutionary schemes. Motivated by Goguen's discussion of blending semiotic structures, we notice a role for dynamical blending in classic works of mathematical philosophy from Lakatos and Pólya. We turn from these inspirations to a survey of analytic frameworks in dialectics, logic, and pragmatics, corresponding to the three layers of IAT. Collating these, we collect 7 sets of tags (of which 6 relate directly to the mathematics domain), which we apply to code several short texts, including student dialogues and online collaborative problem solving exercises, and two short single author works as a comparison case (one of these written by a computer program). This helps to illustrate relationships and interdependencies among the several tag sets.

We build on this work to outline a potential programme for future research in mathematical AI that would draw on contemporary research into social intelligence and mathematical reasoning. A comparison with social creativity in music brings out further issues (indeterminacy, non-verbal communication, problem selection, and context-dependence) that will be relevant to research in computer mathematics and applications of the ideas discussed here to other domains.

# Contents

# 1 Introduction

We will present a catalogue of examples of social creativity in mathematics, together with a survey of existing approaches to their socio-linguistic and "meta-mathematical" interpretation. Music provides a comparison case that helps to flesh out a theory of social creativity – showing the limitations, for example, of an entirely linguistic approach. William Thurston [75] talks about why social creativity is important for mathematics in practice:

> *We are not trying to meet some abstract production quota of definitions, theorems and proofs. The measure of our success is whether what we do enables people to understand and think more clearly and effectively about mathematics.*

Our long-term goal is build up a systematic, and ultimately implementable, computational account of the way mathematical understanding develops. The examples in which we ground our work at this early stage are of three kinds:

(A) Live, in-person mathematical dialogues may show best how mathematical sociality works in practise, although recorded examples are somewhat limited, and tend to contain many extra-mathematical features. These examples are useful for theory development, but harder to systematise.

(B) At another extreme, recent developments in computer mathematics have produced proofs written in a manner that resembles natural language, while other systems provide basic examples of computational social intelligence. We would ultimately like to have a system that can replicate both feature sets.

(C) Finally, perhaps the most central set of examples for this work are an existing catalogue of social experiments known as the MiniPolymath projects. These focus on computer-mediated problem solving activities at the advanced student level. In the future, we should be able to replicate features of these discussions, for example, by running a series of "MiniPolymaths-for-Machines".

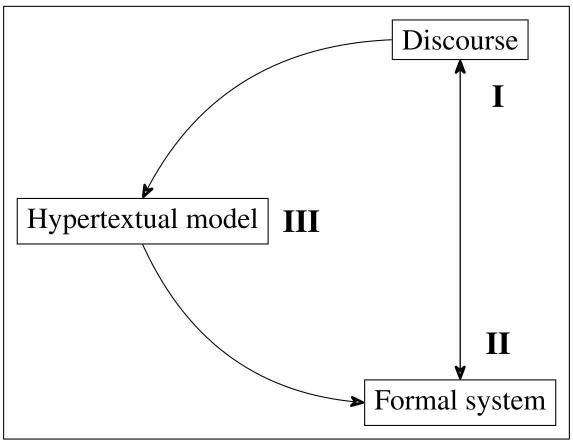


Figure 1: Schematic

The idea underlying our treatment of these examples is to look at how a relatively informal mathematical discussion corresponds, in an appropriate sense, to set of formal and computationally meaningful objects, such as a proof plan [12]. Figure 1 presents a general schematic that depicts this situation, and Figure 2 shows one possible instantiation. We will refer to each of the evolving components as "diagrams", although they need not be expressly pictorial. We envision both the possibility of participants in a discussion directly modifying the state of a running computation (I→II) – and of the computer intervening in the discussion as a participant (II→I).

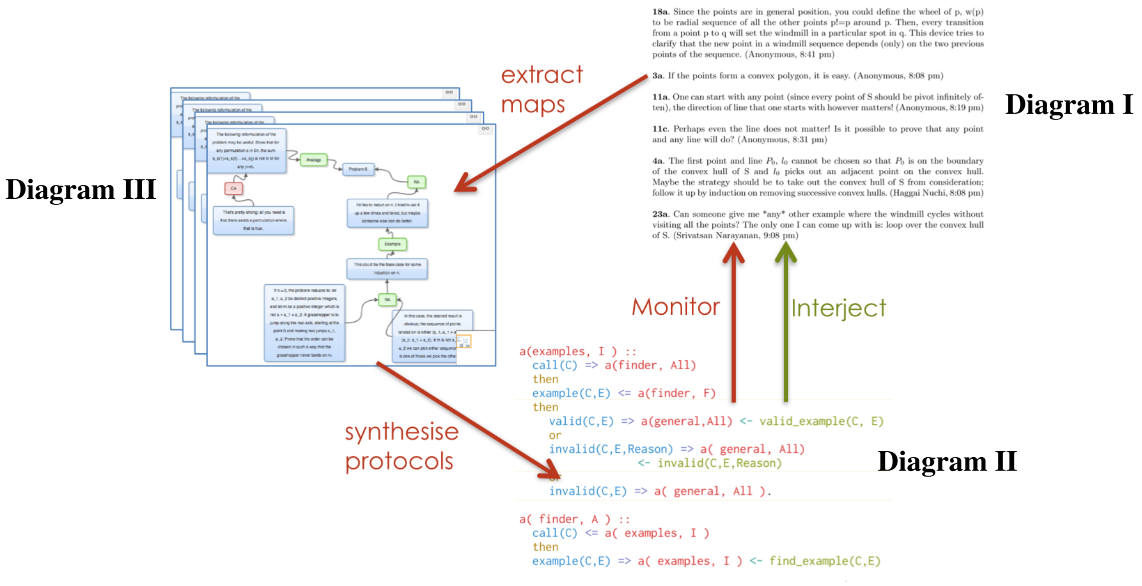


Figure 2: The Online Visualisation of Argument (OVA) system as a bridge between Minipolymath3 and a discourse model in the Lightweight Social Calculus

In the typical case, moves in Diagram I drive moves in Diagram II, mediated by a model in Diagram III. The situation as a whole resembles the general case of *heterogeneous reasoning* as described by Barwise, and instantiated in his Hyperproof system [5]. In Hyperproof, Diagram I would be a graphic with certain implicit rules governing its structure, and Diagram II the corresponding logic problem. In a typical mathematics setting, Diagram I may include some explicit pictorial features, but its basic structure is (implicitly) that of a graph or hypertext that depicts a dialogue. This structure is made explicit in Diagram III. Diagram III has the important feature that it reflects all of the structures in Diagram I that have a recognisable mapping into Diagram II, and, symmetrically, there are no language-level features in Diagram II that do not have an analogue in Diagram III. As Barwise [5] put it: "*Inference, as we understand the term, is the task of extracting information implicit in some explicitly presented information.*"

Our strategy in the paper is to follow the parallels between dialogue and the interpretation of dialogue, logic and models of logic, and contextualised discourse and its hypertextual models. While we our overall approach considers examples in categories A ("live discussion"), B ("computer mathematics"), and C ("computer-mediated discussion") in terms of features from each of Diagram I, II, and III, previous research has not always been so synthetic, and often focuses on one of these components, rather than on their interconnections.

Our overall philosophical perspective on *sociality* derives from that of George Herbert Mead [44], witnessed here by the key feature that both Diagram I and Diagram II are understood to be evolving works-in-progress. If either is completed, the conversation ends, or the proof is resolved. This is similar to the Meadian (multi-)perspective in which both "organism" and "environment" undergo change together. In this metaphor, Diagram III corresponds to the sensory interface between "organism" and "environment." Alternatively, from a linguistic point of view, Diagram III serves as a Rosetta stone that can be used to define translations. More formally, it is the existence of satisfactory mappings between discourse on the one hand and the logical description on the other that allow us to think about connections between the two diagrams in terms of Barwise and Seligman's *infomorphisms* [6]. Barwise and Seligman would call Diagram III the *core* of the information channel.

The work presented here is a preliminary step in a larger programme that aims to identify *dialogical* moves that can be associated with meaningful computational moves.[1] In the mathematics domain, which is our primary focus in this paper, the relevant computations are typically *logical*.[2] Other dualities (e.g. content and expression, genotype and phenotype, etc.) may potentially be treated similarly: the key issue in each case is their *pragmatic* interplay – i.e. the effect of context on the outcome of events. The role of context is well-studied in linguistics and the philosophy of language, since merely decoding statements is not usually sufficient to determine the intended meaning [69]. Context has also been studied in computing, although earlier attempts to formalise the concept are not conclusive [68, 30] – with Hirst [30] in particular arguing that "*the notion of 'context' can be defined only in terms of its effects in a particular situation.*" Our notion of context is embedded in Diagram III. The relationship of this notion to the metaphors of "perception" and "translation" introduced earlier can be conveyed with a simple computer interface example, namely a switch that converts a computer keyboard from the QWERTY to the Dvorak layout. Whereas both keyboard layouts can model the same discourse (as represented by a flow of characters), throwing the switch influences the practical characteristics of the model. In short, Diagram III may be stateful.

Our proposal to describe the pragmatic and computational features of mathematical discourse is similar to earlier work described in [40], where the authors examined a specialised linguistic task[3], and found $n$-grams that were strongly associated with pre-defined speech acts [3, 15]. Here, we present necessary preliminary work, charting a course for future NLP-based analyses. Our current objectives are relatively high-level in comparison to contemporary work on the linguistic analysis of mathematical texts [20, 80]. What we offer here is a synthetic, computational, approach to mathematical understanding, drawing on philosophy of mathematics and adjacent fields.

Our first aim in the current paper is to outline the set of speech acts that underlie mathematical dialogues. Although mathematics is a fairly specialised domain that is communicated with an array of technical languages, as it is used in practise, mathematical language is also tremendously expressive. Classifications of speech acts in expressive languages tend to be correspondingly broad and nuanced [39, 74]. We address this aim through a survey of existing frameworks, and a preliminary application of these frameworks to code several examples of mathematical dialogue at differing levels of formality.

Our second aim is to develop an approach that can replicated to model social creativity in other areas. The claim to generality is related in part to core theories that we build on, as described in Section 2. In the concluding discussion section, we will look at how our approach could apply in the music domain. This serves to draw out some general issues related to indeterminacy, nonverbal communication, problem selection, and further explain the role of context.

[1] Dialogic is more general than *dialectic*, which aims to converge on a common ground [65, pp. 18–20].

[2] Other discourses appeal, for example, not to reason, but to the emotions, or to arguments based on character (*viz.*, *logos*, *pathos*, *ethos*) [13].

[3] http://groups.inf.ed.ac.uk/maptask/

## 2 Background and Related Work

The primary ingredients we will use to build a formal theory of social creativity are Barwise and Seligman's Information Flow [6], Goguen's Unified Concept Theory (UCT) [22, 24], and Lakatos's Proofs and Refutations [37]. In the broader features of the model introduced in Section 1, we are also inspired by the work of Katarzyna Budzynska and Chris Reed on Inference Anchoring Theory (IAT) [58, 10, 9], which develops connections between a dialogical context and logical argument by way of illocutionary schemes. In this manner, IAT takes the earlier notion of *dialogue games* [16] (also known as *dialogue systems*) in a more explicitly computational direction. "[W]*hereas logic defines the conditions under which a proposition is true, dialogue systems define the conditions under which an utterance is appropriate*" [57]. Like the work we present here, IAT aims to build a bridge between dialogue and logic.

In Section 1, we briefly introduced the notions of infomorphisms and information channels. The more formal definition of an information channel revolves around a *core*

$$\mathscr{C} = (\text{tokens}(\mathscr{C}), \text{types}(\mathscr{C}), \models_{\mathscr{C}})$$

where $\models_{\mathscr{C}}$ is a binary relationship between tokens and types; specifically, $t \models_{\mathscr{C}} T$ stands for the statement "$t$ is classified as being of type $T$ in $\mathscr{C}$." The core is complemented by an indexed collections of classifications and maps

$$\{A_i = (\text{tokens}(A_i), \text{types}(A_i), \models_{A_i})\}_{i \in I}$$

$$\{f_i^{\check{}} : \text{tokens}(\mathscr{C}) \to \text{tokens}(A_i), f_i^{\hat{}} : \text{types}(A_i) \to \text{types}(\mathscr{C})\}_{i \in I}$$

with the property that

$$\forall i \in I, c \in \text{tokens}(\mathscr{C}), \alpha \in \text{types}(A_i) : f_i^{\check{}}(c) \models_{A_i} \alpha \text{ iff } c \models_{\mathscr{C}} f_i^{\hat{}}(\alpha)$$

The elements $c \in \text{tokens}(\mathscr{C})$ are called *connections* between the various tokens $f_i^{\check{}}(c)$. A good example would be English sentences (tokens) and their classification as dialogue moves (types, e.g. "attack", "conflict", "question", etc.), which are mapped to into a graphical representation that shows the pattern of dialogical response together with the associated illocutionary points (e.g. assertions, inferences).[4] The result could be placed in correspondence with a computational model of the argument's logical structure.

In our setting, a locution $E$ that presents *an example of a concept* might map to a hypertextual representation that connects the example $E$ to the concept $C$ that it exemplifies, which could then be mapped in a formal model of the dialogue to the statement `example(C,E)`. The Lightweight Social Calculus [48] (also known as the Lightweight Coordination Calculus [60]) mentioned in Figure 2 is one suitable target language. Note that we do not require every conversation to happen within the same scope or the same formal model, since information channels can themselves be chained together and given multiple interpretations. One of the processes with which we will be particularly concerned here is *blending*, in which using two different models, the result is not either/or or both/and, but something more than the sum of the parts.

Blends are present both in the informal logic of proof and the literature on mathematics problem solving. However, these settings tend to bring a dynamical and empirical approach to blends

[4] See [59, 11] for an expanded discussion of this sort of example.

that is not present is straightforward "concept blending" [22, 24]. This is paralleled in Borgo and Goguen's discussion about dynamical systems and creativity in the context of music [8]. Goguen's earlier work on concept blending was not of direct use as a core theory of musical creativity, "*because musical structure is inherently hierarchical, and hence cannot be adequately described using only atomic elements and relation instances among them*" [23]. However, the model was largely salvageable as they turned to analogous *structural blending*, which draws upon the idea of semiotic systems with level-ordering and priority ordering: "*Levels express the whole-part hierarchy of complex signs, whereas priorities express the relative importance of constructors and their arguments.*"

That said, over the last 100 years there have been varied innovations in music (e.g. 12-tone music, free jazz, free improv) with non-hierarchical structures and representations. For our purposes it is worthwhile to note that semiotic structures can also express dynamics through the use of recursion [2], although Goguen did not pursue this. We are interested in *dynamical blending* because of the importance of modelling not just hierarchical content, but evolving, and emergent, work-in-progress. Not only does discourse and its formal representation change over time, the very terms by which this representation is made may change as the context changes.

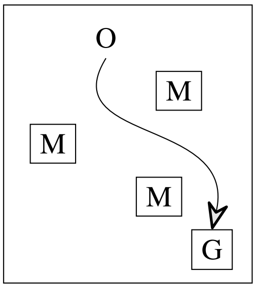

Figure 3: Paths as blends

Figure 3 presents a simple illustration of the idea: a path is traced out as a blend of several forces. Commenting on a similar image, Andersen [2] writes that the basic metaphor for thinking is *travel*. But rather than considering a simple path between obstacles, we might envisage a skier descending amongst moguls. Through continued use, the landscape shifts, and the classifications of paths in terms of their homotopic features or their desirability may change. The "*relations between relations*" [34] that define semiotic systems can be hooked together and react back on themselves, as our representations, relations, and the world we live in evolve over time.

In a mathematical setting, there seem to be blends associated with Lakatosian discourse that take a *conjecture* and an *example* and that, depending on contextual features, yield a revised conjecture, or a revised example, corresponding to an update in a developing theory. Lakatosian dialogue is based on empirical processes, although develops in relation to and with reference to a logical structure. It is through an informal and often hands-on process of conjectures and counterexamples that the underlying formal logical structure of a mathematical domain is made clear. Section 3 includes further discussion of Lakatosian discourse.

In Pólya [53], there is a parallel constructive blend that develops in the course of decomposing a given problem into several interrelated *auxiliary problems*. The blend, indeed, describes these relationships, typically in terms of new unknowns. As the problem solution progresses, these unknowns are determined using the given data. There may be a degree of indeterminacy in the solution strategy (e.g. multiple possible paths to the goal in Figure 3), while any given solution traverses various intermediate states between the given data and the goal as it is worked out. At another level, *problems* (and, indeed *solutions*) can be combined, like the differential operators in quantum mechanics. Gilles Deleuze [19] considers related issues from a broad philosophical perspective[5], drawing particularly on Albert Lautman's earlier discussion of dialectics in mathe-

[5] On auxiliary problems, see esp. [19, p. 239].

matics; see [38] for a relevant survey.

In its ability to "*Find a block which is taller than the one you are holding and put it into the box*" [79], Winograd's SHRDLU already combined several dynamic concepts, albeit in a simple way. More recently, Push Singh [66] studied social collaboration in a construction task with more complexity than anything found in the Blocks World. Singh's program EM-ONE needed hundreds of *narratives* and *mental critics* for the agents to use as they walked through the simulated collaborative construction process. It would not be surprising if we end up needing even more complex heuristic strategies in order to simulate or effectively model mathematical creativity.[6] Singh drew on Minsky's idea of a *panalogy* architecture [67, 46] and we may find that we need a similarly complex multi-layered architecture. Although Singh did not consider blends, it might be natural to look at blending two narratives or two critics in the same way as we might blend two differential operators that define a problem or problem solving process. Considering Singh's use of LISP-based representations, earlier work in the field of Genetic Programming [35, 36] could inspire an evolutionary approach to the construction of such blends. Another related approach with relevance to the work presented here is a previous application of dialogue games to multi-agent systems [49]. This informed the subsequent development off a layered agent model [32] which is reminiscent of both the Singh-Minsky architecture and Goguen's comments on semiotic systems [23], mentioned above.

Gerry Stahl has done extensive research on mathematics in a social context [71, 72, 73], developing research into "*group cognition*," drawing on online interactions between students. His notion of "*adjacency pairs*" is a broader category than what we call Lakatosian moves. Stahl's work focuses on computer supported collaborative learning, which is an inspiring domain for our work, although we also aim to add dimensions related to computer *simulated* collaborative learning. Nevertheless, our approach in the current phase of work, in which we will closely examine real-world dialogues, is similar to Stahl's. Contemporary strategies from natural language processing , including the field of *argument mining* [52] which has gained recently traction within the broader field of *discourse mining* (e.g. [70, 78]) are also closely related, although here we only developing groundwork for future NLP-based efforts. In the current work we will not only look at argumentative structures, but also at pre-argumentative structures, that is, we consider the constructive features of *informal logic* (cf. [31]).

# 3 Survey of analytic frameworks

In this section we will discuss a range of frameworks, under the headings "Diagram I", "Diagram II", and "Diagram III" depending on whether they have more to do with dialogue, logic, or pragmatics, respectively. We note that these analytic frameworks are closely parallel to the three types of examples (dialogical, computational, and computer-mediated) introduced in Section 1. We will turn to concrete examples in Sections 4 and 5. We can justify presenting the analytic frameworks first, since they help to guide our selection of examples.

**I.** Douglas Walton considers the basic patterns in dialogue to be: ***persuasion***, ***negotiation***, ***information seeking***, ***deliberation***, and ***quarrel*** [77]. In some earlier writings these were augmented

[6] Cf. Gowers's Tricki wiki, `http://www.tricki.org`.

with a sixth category *inquiry*, and with the subtype *pedagogy* and mixed form, *debate* [76]. Andrew Aberdein [1] indicates that "*Many other familiar argumentational contexts may be represented in terms of Walton's six basic types of dialogue by such hybridization and subdivision*." In the mathematics context, Aberdein offers a discussion of proof dialogues instantiating Walton's dialogue types. Thus, we have *proof as inquiry*, *proof as persuasion*, *proof as information seeking*, *proof as deliberation*, *proof as negotiation*, *proof as eristic/debate*. Each of these is described in terms of an *initial situation*, the *main goal* the *goal of the prover* and the *goal of the interlocutor*. Depending on which type of proof is under discussion, Aberdein indicates that the *main goals* are:

1. ***Prove or disprove conjecture*** (inquiry),
2. ***Resolve difference of opinion with rigour*** (persuasion),
3. ***Transfer of knowledge*** (information seeking)
4. ***Reach a provisional conclusion*** (deliberation),
5. ***Exchange resources for a provisional conclusion*** (negotiation),
6. ***Reveal deeper conflict*** (eristic).

In general "*an utterance is appropriate if it furthers the goal of the dialogue in which it is made*" [57]. Henry Prakken [57] focuses on *persuasion dialogues*. He explains that in this context, some of the typical dialogue features are: ***arguments*** and ***questions***, ***claims***, ***challenges***, ***conceding***, and ***retraction***.

In the context of informal mathematical discussions, Pease et al. [50] recast persuasion in terms of *Lakatosian moves* relevant to conjectures and proof:

1. A ***conjecture*** is what is argued for.
2. ***Surrender*** consists of abandoning a conjecture in light of a counterexample.
3. ***Piecemeal exclusion*** defends a conjecture by dealing with exceptions through the exclusion a class of counterexamples.
4. ***Strategic withdrawal*** uses positive examples of a conjecture and generalises from these to a class of object, then limiting the domain of the conjecture to this class.
5. ***Monster-barring*** argues that a 'counterexample' can be ignored because it is not a counterexample, as it is not within the claimed concept definition. Using this method, the original conjecture is unchanged, but the meaning of the terms in it may change.
6. ***Monster adjusting*** is similar, in that one reinterprets an object in such a way that it is no longer a counterexample, although in this case the object is still seen as belonging to the domain of the conjecture.
7. ***Lemma incorporation*** distinguishes between *global* and *local* counterexamples. The first would be counterexamples to the main conjecture, and the latter is a counterexample to one of the proof steps or lemmas.

In domain non-specific work on the Speech-Act Annotated Corpus (SPAAC) presented in [39], speech acts are conveniently organised into 5 superordinate categories (***expressive***, ***interpersonal management***, ***dialog control***, ***mainly initiating***, ***mainly responding***), plus several acts classified as "other" (***external to dialog goals***, ***unspecified***).

**II.** As mentioned above in connection with the systems SHRDLU and EM-ONE, it should be possible to model dynamical behaviour by applying "concept blending" to simple operators. The Lakatosian moves just discussed can be thought of as operators whose outcome depends on the way in which a new example is construed:

$$\text{conjecture} \oplus \text{counterexample} \mapsto \text{surrender this conjecture} \tag{1}$$
$$\text{conjecture} \oplus \text{exception} \mapsto \text{exclude of a range of examples} \tag{2}$$
$$\text{conjecture} \oplus \text{examples} \mapsto \text{limit scope} \tag{3}$$
$$\text{conjecture} \oplus \text{monster} \mapsto \text{revised statement} \tag{4}$$
$$\text{conjecture} \oplus \text{monster} \mapsto \text{revised example} \tag{5}$$
$$\text{conjecture} \oplus \text{counterexample} \mapsto \text{revised approach} \tag{6}$$

In particular, the "revised approach" in (6) may be associated with additional lemmas and changes to existing lemmas or to the conjecture. There is no reason why (1)–(6) should limit the range of possible moves. Thus, one would expect that there to be a set of discourse moves that take place at the "managerial" level and that move the discourse between conjectures in ways that are not explicitly connected with (counter)examples. There are many other "computational" heuristics that may underlie and motivate the discourse moves mentioned above.[7] Pease and Martin [51] introduced a related typology of comments that move a mathematical discussion forward: those pertaining to ***concepts***, ***examples***, ***conjectures***, ***proof***, and ***other***.

These Lakatos-inspired reflections remain somewhat abstract in the sense that they deal with mathematical conjectures, examples, and lemmas but do not consider other kinds of (often entirely mundane) mathematical objects. Corneli [18] describes the following kinds of mathematical object types that are either supported in PlanetMath.org or planned for a future version of the PlanetMath/Planetary system: ***article*** [$A$], ***link*** [$\ell$], ***project*** [$\mathbb{X}$], ***post*** [$T$], ***solution*** [$S$], ***review*** [$R$], ***update*** [$\sharp$], ***question*** [$Q$], ***correction*** [$C$], ***fork*** [$/$], ***outcome*** [$\star$], ***problem*** [$P$], ***collection*** [$L$], ***classification*** [$M$], ***conjecture*** [$\mathbb{J}$], ***group*** [$G$], ***user*** [$U$], ***request*** [$W$], ***heuristic*** [$H$], and ***ephemera*** [$\mathbb{E}$]. On PlanetMath, the underlying logic depends on implementation features that describe how these elements connect and interact. In everyday language, the various objects can be directly combined with comments, like those described just above, for example, in a conjecture like "the ideas in this article suggest the following..."

Whether the development of a mathematical theory is thought of in terms of dynamically constructed structure blends or growing networks of interrelated terms, within Diagram II, we require a range of more or less mechanical operations to carry out the reasoning steps. One recent computational example has been developed by Ganesalingam and Gowers, who describe an automatic problem solving program ROBOTONE [21] that, they claim, works roughly in the same way that people do. While this claim may be debated, programs that do not come with a similar claim would not provide a schematic mapping between "human operations" and "machine operations." As our initial goal in the current work is to understand human operations in a social context, it is helpful to have a potential target language to map into, even if it ends up being a moving target. In this system, "*An individual move is an operation that transforms a specific problem state into another*

[7] For example, Minsky [47] mentions: *reasoning by analogy*, *dividing the problem into parts*, *changing the problem's description*, *focusing on a more specific example*, *making a simpler version*, *trying to identify what makes the problem hard*, *imagining what an expert would do*, *stopping what doing if you're stuck*, *retrieving the knowledge that tells you how to solve the problem*, and *asking for help*.

*state in a sound fashion; thus individual moves correspond to application of tactics to a specific LCF-style prover state.*"[8] Ganesalingam and Gowers describe, in total, 27 different moves, from move categories ***Deletion***, ***Tidying***, ***Applying***, ***Suspension***, and ***EqualitySubstitution***. The basic heuristic is to look at the problem state and apply the first possible move from an ordered list.[9]

**III.** The way people solve problems has been studied extensively, with the best-known work being that of George Pólya (e.g. [54, 55, 56, 53]). Empirical work continuing in this tradition has been developed by Alan Schoenfeld (e.g. [62, 63, 64]) and others.

Pólya's heuristics follow the well-known outline,

1. ***Understand the problem (unknown, data, condition)***,
2. ***Find the connection between the data and the unknown to obtain a plan of the solution***,
3. ***Carry out the plan***,
4. ***Examine the solution obtained***.

The empirical aspects of Schoenfeld's work connect with earlier protocol-based methods for doing research in mathematics education, e.g. Lucas [41], and Lucas et al. [42] who built on earlier work in process coding by Jeremy Kilpatrick [33] as well as early work by Schoenfeld himself (cf. [61]). However, Schoenfeld found that the codings used in [42] to be overly complex, while still focused only on tactics rather than strategy. Schoenfeld insists that *omissions* should be coded for, along with explicit behaviour. His primary coding is divided among stages that are clearly informed by Pólya: ***Read***, ***Analyze***, ***Explore***, ***Plan***, ***Implement***, and ***Verify***. These are supplemented by ***New information and local assessments***, and ***Transition***. In studying problem solving, Schoenfeld argues "*it is what the person does rather than what the person produces*" that is important [62, p. 4]. One of Schoenfeld's central research strategies was to put people in small groups or pairs and have them talk through mathematical problems together. He points to two reasons for this choice: one is to alleviate situational pressure. "[W]*hen two students worked on the problem it was typical for one student to turn to the other and say something like, 'I have no idea what to do. Do you?'*" This not only produced more natural records of a thought process, but also helped to avoid answers that were formal for the sake of being formal [62, p. 279–281]. Nevertheless, a social setting is no guarantee of success, as becomes clear from several dialogues that record attempts to solve this problem:

> *Three points are chosen on the circumference of a circle of radius R, and the triangle containing them is drawn. What choice of points results in the triangle with the largest possible area? Justify your answer as best you can.* [62, p. 319]

After a brief period of time spent *reading*, the first group of students spends the rest of the time *exploring* in a rather meandering fashion, even though they had a suitable conjecture within the first few moments of discussion [62, p. 294]. Schoenfeld remarks [62, pp. 288–289]: (1) The students

[8] LCF stands for *Logic for Computable Functions* (cf. [45, 26, 25]).

[9] Gordon [25] explains why activities related to *deletion* are seen to have particular priority in this type of program: "*The steps of a proof would be performed but not recorded, like a mathematics lecturer using a small blackboard who rubs out earlier parts of proofs to make space for later ones.*"

neglected to assess the potential utility of calculating the area of the equilateral triangle; (2) the discussants gave consideration to several interesting alternative problems that might have added insight, but did not pursue them vigorously; (3) progress was not monitored or assessed during the solution. A second group described a "hunch" that the answer was an equilateral triangle, but spent most of the available time *implementing* a somewhat related demonstration to show the size of the largest inscribed *right triangle*. They had no satisfactory answer when asked by the investigator how what they found would relate back to the original problem. In short, as Schoenfeld observes: "*If one makes major strategic mistakes, then matters of tactics are of little importance*" [62, p. 289].[10]

In this connection, the evolution of Pólya's framework in [53] is relevant. Here, he particularly focuses on *planning*.

1. ***Identify the goal*** (what do you want?)
2. ***Identify the conditions*** (what do you have?)
3. ***Decompose the problem*** (introducing ancillary problems and new unknowns as needed)
4. ***Connect the data to the problem***
5. ***Determine any unknowns***

Schoenfeld introduces a "macro-level" framework [62, p. 15] that supplements the "micro-level" frameworks derived from Pólya, and that describes thinking in terms of four categories: ***resources*** ("genetic epistemology"), ***heuristics*** (per Pólya), ***control*** ("decision making" and "metacognition"), and ***belief systems*** (or "mathematical world view"). He points out that this framework is "*far from comprehensive*" and that it generally excludes both cognitive and social details [62, p. 16]. If we consider a social implementation, the issues are reversed, and we must find a way to embed features like heuristics, control structures, and belief systems within a social rather than personal context.

Pease, Aberdein, and Martin (forthcoming) are working on *explanation* in mathematical texts. Explanations can be people-centred (paralleling Schoenfeld's macro-level framework) or domain-centred. Note that items on the following list can function either as an *explanandum* or an *explanans*. For example, we might attempt to explain why we don't understand something, or we could point to the fact that we don't understand something *as* an explanation.

1. ***abilities*** (what can/can't we do, e.g. we can reduce the problem to $P$)
2. ***knowledge*** (what do/don't we know, e.g. $X$ is wrong)
3. ***understand*** (what do/don't we understand, e.g. do you see why this is a contradiction?)
4. ***value/goals*** (what do/don't we want, e.g. $X$ is a good idea)
5. ***initial problem*** (e.g. the initial problem is harder if $P$)
6. ***proof*** (e.g. $A$ is not a useful approach)
7. ***assertions*** (e.g. $M$ is subset of $P$)

[10] "Tactics-based proof systems" may or may not concern themselves with strategic matters; strategy does not appear to play a particularly significant role in the system described by Ganesalingam and Gowers [21].

8. ***specific cases/instances*** (e.g. there will always exist instance $X$ that satisfies condition $C_1$
9. ***arguments*** (e.g. let us suppose $X$. Then $Y$.)
10. ***representation*** (e.g. by reducing the problem to $P$)
11. ***property*** (e.g. We don't know if it has property $P$)

# 4 Methods

Referring to the foregoing material, we could draw from a large array of possible annotations at different levels of specificity (Table 1). Clearly there is redundancy among the elements of this list, as we have attempted to illustrate, initially, with the numbering scheme (*1a.*, *1b.*, etc.). However, even taking this level of redundancy into account, a total of 7 possible dimensions for tagging (of which 6 relate directly to the mathematics domain) put us a position at least as complex as that taken up by Lucas et al. [42] – indeed more complex since these authors considered only individual problem solvers working with a speak-aloud protocol. At this point in our study, it seems best to take a comprehensive view than to risk leaving something out for the sake of concision. Accordingly, in this design-oriented phase of work, we will intensively annotate several brief passages, drawing as needed from all of the above formats, as we begin to develop a suitable apparatus. In subsequent phases, when the coding and technologies have stabilised, we will deal with longer passages (including full proofs), and will report on features like intercoder reliability. In the current study, we instead sought direct consensus for coding of the sample texts. In Section 6, we will compare the resulting codes with those used in [42].

**Data sets.** We are particularly interested in the MiniPolymath problems that have been posed, discussed, and solved on Terrence Tao's blog and the Polymath blog. There have been four such discussions to date; however, the experiment would be relatively easy to replicate. The problems – drawn from Mathematics Olympiads – are interesting in that they are challenging enough to spur considerable discussion, but not so challenging as to go unsolved when people put their heads together.[11] They do not yet have the scope of the full Polymath projects, which deal with research topics (cf. [4, 28]), which we hope to tackle in future work. Unlike Lakatos's *Proofs and Refutations* [37] – which remains inspiring – the data is "real" as opposed to "reconstructed in fictional form." The MiniPolymath data is also publicly available online. The questions that were discussed in the four MiniPolymath sessions that have taken place to date are presented in an Appendix. In the current paper we also consider brief excerpts from Schoenfeld's data set (several dialogues between students are presented in full in his book), an excerpt from a monologue presented in a public lecture by Timothy Gowers on joint work with Mohan Ganesalingam on mathematical discovery [27], and a proof written by a prototype system named ROBOTONE developed by Ganesalingam and Gowers [21] that operationalises the line of thinking from the lecture.[12] Single author works addressed to an arbitrary audience *can* be coded using the frameworks we've described, although this is not in general the intended application. Single author works provide a natural point of comparison, as we think about what the "social dimension" brings to mathematics.

[11] Tao notes, regarding MiniPolymath1: "*Of the 500-odd participants in the Olympiad, only a half-dozen or so managed to solve this problem completely.*"

[12] http://people.ds.cam.ac.uk/mg262/robotone.pdf

| | | |
|---|---|---|
| *0.* | The general-purpose *SPAAC Classified List of Speech Acts* | (“SA”) |
| *1a.* | Walton’s *patterns of dialogue* | (“WD”) |
| *1b.* | Aberdein’s interpretation of these in the context of *proof dialogue* | (“AD”) |
| *2a.* | Prakken’s specialisation to *persuasion dialogues* | (“PD”) |
| *2b.* | Pease et al.’s approximately parallel *Lakatosian moves* | (“LD”) |
| *3a.* | Pólya’s *stages of problem solving from “How to Solve It”* | (“PS”) |
| *3b.* | Schoenfeld’s refinements to this in the form of his *process coding* | (“SS”) |
| *3c.* | Pólya’s *stages of planning from “Mathematical Discovery”* | (“PP”) |
| *4a.* | Schoenfeld’s framework describing *factors in mathematical thinking* | (“SF”) |
| *4b.* | Pease, Aberdein, and Martin’s *components of explanation* | (“CE”) |
| *5a.* | PlanetMath/Planetary’s *types of mathematical objects* | (“MO”) |
| *5b.* | Pease and Martin’s *types of mathematical comments* | (“MC”) |
| *6.* | Ganesalingam and Gowers’s *LCF-style tactics* | (“RO”) |

Table 1: List of frameworks, with two-letter abbreviations

# 5 Results

Several short texts have been marked up with codes corresponding to the frameworks introduced in Section 3. The coverage from the several tag-sets is illustrated in Figure 4. Tags attached to dialogical and monological texts have been distinguished, with contributions from single-author texts added as an “increment” above the tags used in dialogues. Table 2 collects the tags that were used, divided into three segments depending on whether the tag set primarily describes (I) dialogical, (II) logical, or (III) pragmatic discursive manoeuvres (see Section 3). The annotation “*” means that all of the tags in a given tag set were used, up to redundancy among the various tag sets. Tags that are redundant in this sense are enclosed in hard brackets in the category to which they were associated in our tagging. The annotation “°” indicates that all but one of the tags was used: namely, *plan* from the SS tag set is not used (while a detailed set of tags related to planning from the PP tag set was used fairly extensively).

Figures 5–9, respectively, reproduce our tagging of: a short passage from a student problem solving dialogue, a single-threaded discussion in Minipolymath 1, a portion of one of the threads in the discussion in Minipolymath 3, the full solution to a challenge problem presented by mathematician Timothy Gowers, and a portion of a textbook-style problem solved by Ganesalingam and Gowers’s program ROBOTONE. Although they are recorded, we did not tally the tactical moves produced by ROBOTONE (RO). The coding in Figure 9 corresponds to mainstream practices of “expanding the definition.” We also did not include Walton’s patterns (WD) which in this context would be redundant with Aberdein’s interpretation (AD); and similarly, we do not include tags denoting Pólya’s stages of problem solving (PS), preferring Schoenfeld’s slightly more detailed rendering (SS).

The most popular tag set was Schoenfeld’s “macro-level” framework (SF), with 11 applications overall. Many of these deal with the application of specific heuristics, such as: *if you can compute something, do it!*; *it’s a good idea to decompose a problem into sub problems*; and *try a simple case*. Note that a discussion about a heuristic, as in Lines 14-16 of Figure 5 – tagged “*heuristic (decompose)*” – is treated differently from the application of a heuristic, as in Lines

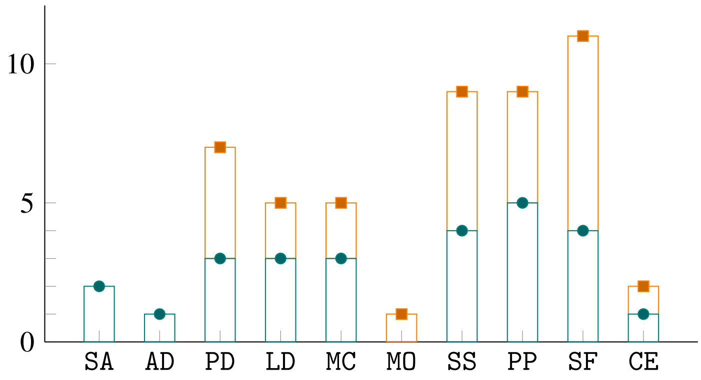


Figure 4: Count of tags from the several schemes

| | |
|---|---|
| `SA` | ackn, inform |
| `AD` | negotiation |
| `PD` | question |
| `LD` | [conjecture], lemma incorporation |
| `MC*` | concept, example (arbitrary inst.), example (monster), [proof], other (phatic) |
| `MO` | ephemera |
| `SS°` | read, analyze, explore, implement, verify, local assessments, transition |
| `PP` | goal, conditions, decomposition |
| `SF*` | resources, heuristic (compute!), heuristic (decompose), heuristic (formal gen.), heuristic (simplify), heuristic (symmetry), heuristic (total stuckness), control, belief systems |
| `CE` | property, [assertion] |

Table 2: Summary of tags used, by category

10-17 of Figure 6 – tagged as "*decomposition*" from the Pólya planning (`PP`) tag set.

The "Pólya-Schoenfeld" tag set (`SS`) and Pólya planning (`PP`) were both used frequently, with 9 applications each, although not all of the tags were used. The Lakatos (`LD`) and comment-type (`MC`) tag sets were used with moderate frequency, with 5 applications each. Note that *conjectures* are mentioned in several different tag sets, and for consistency, tagging into `LD` was preferred. We have tagged *examples*, including counterexamples, into `MC`. We only used two tags from the Classified List of Speech Acts (`SA`), despite this being a general-purpose lexicon, and having the largest set of available tags. This sort of general purpose conversational move seems particularly likely to happen in in-person, real-time dialogues, which are not well represented in our data set.

We included only one tag, *ephemera*, selected from our list of types of mathematical objects (`MO`). This appears to be related to the issue of "use versus mention". Thus, for example, discussants typically referred to concepts, rather than to articles where the definitions of these concepts could be found. We also include only one tag, *negotiation*, from Aberdein's interpretation of proof dialogues (`AD`). Aberdein's discussion of the goals associated with proof includes "prove or disprove the conjecture" (corresponding to the tag *inquiry*) as one of 6 different intentional states. In the examples we looked at, this was generally a goal that all of the participants shared, so using this tag would convey no information. We interpreted Lines 13-15 of Figure 7 to be (part of) a brief negotiation about the need for a "backward" operation in an induction proof. Finally, we include one tag, *question*, from Prakken's framework (`PD`).[13] Lines 1-2 of Figure 7, marked as *assertion* using the components of explanation framework (`CE`), might have, synonymously, been marked as a *claim*. The tag "*assertion*" is therefor recorded in hard brackets in Table 2. In addition to directly redundant tags, tagging these sample texts also makes more clear the relationships and interdependencies among the dimensions. For example, as noted above, `PP` supplies a dimension that expands on the idea of *planning* that is noted abstractly in `SS`. `MC` records examples, which are a necessary input for most of the elements of `LD`. The tag sets are also associated with different senses: not only "use versus mention", but also use-to-explain or use-to-question.

# 6 Discussion

## 6.1 General comments

In terms of our earlier decomposition: `SA`, `AD`, `PD`, and `LD` were initially presented in connection with dialogical features; `MC` and `MO`, together with an interpretation of `LD`, in connection with logical features; and `SS`, `PP`, `SF`, and `CE` in connection with pragmatic features. Further work on an implementation layer, as well as an expanded corpus of tagged texts, will be needed to see more clearly how these moves contribute to the development of a proof-in-progress.

Frameworks that are only used for one (`AD`, `PD`, and `MO`) or two (`CE`, `LD`) tags are nevertheless interesting and worthy of further attention and potential refinement. Other terms from these frameworks (for example, Prakken's *challenges*) are likely to come up if we examine longer texts and other settings. Similarly, more instances of *negotiation* and *deliberation* about sub-problems and

[13]For reasons of typographical convenience, the question tag is not explicitly recorded in Figures 5–9 – rather, we count each paragraph that includes at least one question mark as a question. We have also counted the question-like sentence on Lines 13-15 of Figure 7 which is punctuated with a period as a question.

| | |
|---|---|
| DK: [reads the question][ss1] | [ss1] read |
| BM: Do we need calculus for this?[sf1] So we can | [sf1] resources |
| minimize, or rather maximize it.[mc1] | [mc1] concept |
| DK: My guess would be more like [indiscernable] | |
| my basic hunch would be that it would be –[ld1] | [ld1] conjecture |
| BM: An equilateral –[ld1] | [ld1] conjecture |
| DK: 60, 60, 60[mc1] | [mc1] concept |
| BM: Yeah.[sa2] | [sa2] ackn |
| DK: So what choice of points has to be where on | |
| the triangle[pp1] – these points are gonna be.[pp2] | [pp1] goal [pp2] conditions |
| BM: Try doing it with calculus – see if you can – just | |
| draw the circle[ss3] – see what we'll do is figure | [ss3] explore |
| out the right triangle –[ss8] | [ss8] transition |
| DK: Yeah, or why don't we find – or why don't | |
| we know the – some way to break this problem | |
| down into[sf2] – like, what would a triangle be for | [sf2] heuristic (decompose) |
| half the circle?[sf2] | [sf2] heuristic (symmetry) |

Figure 5: Excerpt from Schoenfeld [pp. 324–325]

| | |
|---|---|
| Nate: Well, my first thought is to see if the hy- | |
| potheses seem reasonable.[ss3] The hypothesis that | [ss3] explore |
| $s = a_1 + \ldots + a_n$ not lie in $M$ is certainly neces- | |
| sary, as the last jump that the grasshopper takes | |
| will land on $s$.[pp2] The grasshopper's other steps | [pp2] conditions |
| will land on a partial sums $a_{\sigma(1)} + \ldots + a_{\sigma(k)}$ for | |
| some permutation $\sigma$, but we get to choose the | |
| permutation. Thus it seems plausible that we can | |
| avoid a given set of $n-1$ points.[ss2] | [ss2] analyze |
| Thomas: Quick observation.[sa2] The grasshopper | [sa2] inform |
| must make a first step.[sf2] This is always possi- | [sf2] heuristic (simplify) |
| ble, since the $a_i$ are distinct and $\|M\| = n-1$; that | |
| is, there is always an $a_i$ not in $M$.[pp2] However, | [pp2] conditions |
| let's say $M$ matches all but one of the $a_i$. Then the | |
| first step is uniquely determined. Still, according | |
| to the claimed theorem, a second step must still | |
| be possible.[pp3] | [pp3] decomposition |

Figure 6: Excerpt from MiniPolymath1

| | |
|---|---|
| Haggai Nuchi: The first point and line $P_0$, $l_0$ cannot be chosen[ce7] | [ce7] assertion |
| 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.[mc2] | [mc2] example (monster) |
| 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.[ld7] | [ld7] lemma incorporation |
| Haggai Nuchi: More specifically, remove the subset of $S$ which forms the convex hull to get $S_1$; remove the new convex hull to get $S_2$, and repeat until $S_n$ is convex. Maybe a point of $S_n$ is a good place to start.[ss5] | [ss5] implement |
| Srivatsan Narayanan: Can we just assume by induction that we have proved the result for all the "inner points" $S_2 \cup S_3 \cdot \cup S_n$.[ad5] | [ad5] negotiation |
| The base case would be that $S = S_1$, i.e., it forms a convex polygon.[mc4] | [mc4] proof |

Figure 7: Excerpt from MiniPolymath3

| | |
|---|---|
| Gowers: What is the 500th digit of $(\sqrt{2}+\sqrt{3})^{2012}$?[ss1] | [ss1] read |
| Even this, eventually, a computer will be able to solve.[mc5] | [mc5] other (phatic) |
| For now, notice that total stuckness can make you do desperate things.[sf2] | [sf2] heuristic (total stuckness) |
| Furthermore, knowing the origin of the problem suggests good things to try. The fact that it is set as a problem is a huge clue.[sf4] | [sf4] belief systems |
| Can we do this for $(x+y)$?[ss3] | [ss3] explore |
| For $e$? Rationals with small denominator?[sf2] | [sf2] heuristic (compute!) |
| And how about small perturbations of these?[sf3] | [sf3] control |
| Maybe it is close to a rational?[ld1] | [ld1] conjecture |
| $m^{\text{th}}$ digit of $(\sqrt{2}+\sqrt{3})^n$?[sf2] | [sf2] heuristic (formal gen.) |
| $(\sqrt{2}+\sqrt{3})^2$?[sf2] | [sf2] heuristic (simplify) |
| $(2+2\sqrt{2}\sqrt{3}+3)$[mo20] | [mo20] ephemera |
| $(\sqrt{2}+\sqrt{3})^2+(\sqrt{3}-\sqrt{2})^2=10$[sf2] | [sf2] heuristic (compute!) |
| $(\sqrt{2}+\sqrt{3})^{2012}+(\sqrt{3}-\sqrt{2})^{2012}$ is an integer![ce11] | [ce11] property |
| And $(\sqrt{3}-\sqrt{2})^{2012}$ is a very small number.[ss7] | [ss7] local assessments |
| Maybe the final answer is "9"?[ld1] | [ld1] conjecture |
| We need to check whether it's small enough.[ss5] | [ss5] implement |
| $(\sqrt{3}-\sqrt{2})^{2012} < (\frac{1}{2})^{2012} = ((\frac{1}{2})^4)^{503} = (\frac{1}{16})^{503} < .1^{503}$, so we're in luck...[ss6] | [ss6] verify |

Figure 8: Partial transcript from Timothy Gowers's Maxwell Institute Lecture, November 2, 2012

| | |
|---|---|
| ROBOTONE: If $A$ and $B$ are open sets, then $A \cap B$ is open.[ss1] | [ss1]read |
| Let $x$ be an element of $A \cap B$.[mc2] | [mc2]example (arbitrary inst.) |
| Since $x \in A \cap B, x \in A$ and $x \in B$.[pp3] | [pp3]decomposition |
| Since $A$ is open and $x \in A$, there exists $\eta > 0$ such that $u \in A$ whenever $d(x, u) < \eta$.[pp2] | [pp2]conditions |
| Since $B$ is open and $x \in B$, there exists $\theta > 0$ such that $v \in B$ whenever $d(x, v) < \theta$.[pp2] | [pp2]conditions |
| We would like to find $\delta > 0$ s.t. $y \in A \cap B$ whenever $d(x, y) < \delta$.[pp1] | [pp1]goal |

Figure 9: Partial transcript of ROBOTONE's proof that the intersection of two open sets in a metric space is open

proof strategies would be likely in more wide-ranging discussions (like the research-level Polymath projects). Note, as well, that a framework that has proved to be particularly popular source of tags may be calling out to be separated into several different interrelated frameworks.

Many "standard" problem solving approaches can be implemented in a "social" way (cf. [71], [18, p. 193]). This is hardly surprising in light of the Meadian hypothesis – in which the very emergence of agency is seen as social [43].[14] Although our sample size is too small to make strong conclusions, it nevertheless offer several interesting clues as to what "social thinking" brings to mathematics. In the first instance, as Tao noted the MiniPolymath problems are not easy, and yet, working as a group, participants have been able to solve them at a (time-wise) competitive rate. Why might that be?

People in discussions are presumably more likely to *negotiate*. Indeed, the entire Lakatosian framework might be considered in terms of negotiation and an exchange of resources, rather than simply persuasion and the resolution of differences. Relatedly, discussants may be more likely to *plan*. They are, certainly, more likely to discuss, and as a result may possibly consider a wider range of *examples*. That said, at least in the texts we studied, single authors were more likely to ask *questions*, even if this was only to create a simulated conversation. If they know what they're doing, they may be more likely to simply get on with it and do *problem solving*. They may tend to rely on *heuristics* rather than negotiation to overcome difficulties.

Compared with the 42 primary tags used in [42], the 27 tags that we used are less focused on problem solving *per se*, although there is a significant overlap in the terminology and sources used. The problem solving aspects may be addressable using our combined "pragmatic" tag set (that is, SS, PP, SF, and CE, the segment of our tag collection which bears the closest similarity to [42]) as the core of an information channel, and using these elements to map between "subjective" dialogical moves, and "objective" state changes in a logical setting. However, this set of tags would presumably to be expanded in order to deal with problem identification, positing, and selection.

With ROBOTONE, we get the logical mapping for free, albeit in a monological rather than dialogical format; thus, as they appear in Figure 9, *example (arbitrary inst.)* corresponds to the

[14] "[I]*t is only by acting toward ourselves as others do ... that we recognize and understand ourselves as objects and authors of our own activity. As we learn to coordinate our acting with the acting of others, we differentiate and develop our selves and our abilities to self-determine. Eventually, we not only understand the perspectives (i.e., action orientations and possibilities) of numerous particular others, but also those perspectives explicit and implicit within the broader, more generalized social, cultural practices in which we are immersed*" [43].

moves `expandPreUniversalTarget` followed by `peelBareUniversalTarget`, and *decomposition* corresponds to the move `expandPreExistentialHypothesis`; etc. While ROBOTONE seems to be becoming reasonably proficient at basic problem solving, it currently has no social intelligence. It does not seem unreasonable to expect this system or one like it to expand to include many moves described in the Lucas et al. tag set – without yet requiring social intelligence. After all, as we mentioned above, this earlier tag set had been designed to analyse single-author texts. Working from the other direction, systems like Singh's EM-ONE do possess at least a modicum of social intelligence. It would be natural to extend EM-ONE with narratives related to the `SA`, `AD`, and `PD` (dialogical) tag sets. It would be a bigger challenge to develop the domain-specific knowledge base related to `MC`, `MO`, and `LD`, and mappings between this register and the others.

## 6.2 Extensions to other domains

Naturally, in different domains it will make sense to consider different types of annotations. There are different "relations between relations" to be considered, both within disparate fields of human endeavour, and in different situations within a given field. Nevertheless, analogues to musical experience may be found readily in other domains; music makes a useful initial point of comparison.

Cook [17] proposes to think of a musical score as a "script," rather than as text that is complete in itself. In this way, the space between score and performance is understood as an open and non-linear one, and perhaps more easily conceivable as a continuous exchange between processes and products, comparable to our comments above on the relationship between discussion and proof. Bruce Ellis Benson's work on improvisation-as-dialogue identifies 11 types of performance situations that could cover virtually any type of music [7, pp. 26-30]. Benson's typology ranges from fully notated works (types 1 to 3) to the deliberate subversion of expectations associated with a particular compositional or performative tradition, via real-time improvisation (type 11).

Types 7 to 11 in Benson's typology are well illustrated by examples in jazz and freer forms of improvisation, and are particularly useful in investigating open problem-solving spaces, where structure and meaning are formed and communicated in real time. On that front, Borgo and Goguen [8] attempted a typology of real-time transitions in their joint work on free jazz, in an analysis of a recording by Sam Rivers[15].

1. ***sudden/unexpected segue*** (an unprepared, immediate change with unexpected continuation)
2. ***pseudo-cadential segue*** (an implied cadence with sudden and unexpected continuation)
3. ***climactic segue*** (a peak moment that stimulates unexpected change and continuation)
4. ***feature overlap*** (one feature of the antecedent section is sustained and becomes part of the consequent section)
5. ***feature change*** (a gradual change of one feature that redirects the flow, usually subtly)
6. ***fragmentation*** (a gradual breaking up, or fragmenting, of the general texture and/or rhythm)
7. ***internal cadence*** (a prepared cadence followed by a short silence then continuation with new material)

[15] This work would match types 10 and 11 in Benson's categories.

In their experimental work with free improvisers, Canonne and Garnier [14] relied on musicians' own subjective accounts of short improvised sessions in 3-person teams, and found significant evidence of intersubjective structural patterns emerging in the interactions between participants. This led to a distinction between stable and oscillating behaviours, corresponding to coordinated musical sequences (representable as fixed points in phase space) and discoordinated sections (representable as spaces without fixed attractors). Canonne and Garnier identified four main strategies that improvisers used to convey intentions while playing: ***Stabilization***, ***"Wait and see"***, ***"Playing along"***, ***Densification***. The decision to employ these strategies, as well as the end-result of each strategy employed, depends on the overall musical situation, understanding of other musiciansâ individual intentions and objectives, and on team preferences, based on e.g. competence, range of instruments used, etc. As a result, "misrepresentations" and, by extension, "contrasting evaluations of a given situation" may occur. Improvisers therefore often resort to "meta-pragmatic" evaluations of their strategies while playing, e.g. repeating a strategy until intention is successfully conveyed, or re-evaluating a strategic goal in real-time on the basis of new information gathered from another player's response [14, pp. 202–203]. Similar issues, and more, would apply in any setting where people are not solving predetermined problems.

There are some additional challenges and limitations in the above annotation paradigms. Firstly, given the largely non-discursive, or at least non-verbal, nature of musical performance, any dialogic pattern reflecting decisions vis-à-vis structure or other parameters is usually represented post-hoc, and mostly through the aid of recordings. Other domains (including mathematics) have many non-verbal features that can limit the range of applicability of text-based methods. Secondly, musical situations that correspond to Benson's types 1-6 involve a higher level of consensus as to what constitutes musical meaning, and what formal or structural prerequisites are necessary for its production. They are thus closer to Hall's [29] definition of "low-context" (LC) situations – whereas types 8-11 involve highly subjective, "high-context" (HC) processes. The latter are prone to contrasting evaluations, and pose a range of challenges for modelling. Contexts that change over time or depending on the observer's perspective suggest the need to develop a theory of information flow that takes on emergent properties.

## 6.3 Future Work

The Wordpress blogs that served as a platform for the MiniPolymath experiments, though adequate, do not seem especially *elegant*, and it would be interesting to pursue experiments with other systems that are more tuned to the domain and/or the discourse structure. For now there is plenty of data to be had from extant sources, however, as the understanding of discursive structure discussed here evolves, more of what we learn can be pushed into systems like Planetary [18] made directly available to users, rather than added in a separate secondary analysis.

Developing a version of MiniPolymath for machines would be an interesting related challenge. In this regard both ROBOTONE and EM-ONE are inspiring examples, although neither system would currently be ready to participate in MiniPolymath-like activities, for different reasons. Concerning infrastructure and overall orientation, the Arguing Agents Competition [81] is one point of inspiration, although we prefer a more collaborative tack.

The Graphic Score/Harmonic Spaces case study conducted at A.U.TH. in May-June 2014 qualitatively investigated 11 students' ways of creating a new piece of improvised music while dealing with a combination of two modes of representation: a graphic score open to interpretation, and a series of harmonic spaces drawn from automatic harmonisation trials of early-stage technologies from the COINVENT project (comprised of low-context elements, of Benson's types 1-6, in comparison to the high-context graphic score).[16] A preliminary analysis reveals shared patterns in participants' communication strategies including the following:

1. ***Narrowing the problem space*** (e.g. from an open "what if..." or a more case-specific "what to do with these two sources" to a directional "how can we use source 1 [the harmonic spaces] to interpret source 2 [the graphic score]" and "how do we make this work?")
2. ***Assigning functions and/or meaning to the set material*** (e.g. using particular elements in the graphic score as durational markers, or assigning narrative significance to particular harmonies)
3. ***Mapping sonic elements onto visual ones, and vice versa*** (e.g. creating subscores and testing them via different realisations)
4. ***Defining end-product ontologies*** (agreeing on what the resultant piece should be described as, and what its constituent elements are)

Decisions reached verbally were often reconfigured and in some cases entirely reversed based on playing and listening sessions between discussions. More work is necessary to arrive at a more conclusive representation of how individual aesthetic preferences were negotiated by each group. A more detailed analysis of recorded conversations, performance sessions and reports collected during the study may provide further examples of common or shared strategies employed among participants. In parallel, we plan to pursue extensions to the tag sets used here, and greater alignment in the descriptions of collaborative dynamics in the mathematics and music domains.

## References


[1] ABERDEIN, A. The informal logic of mathematical proof. In *Perspectives on Mathematical Practices: Bringing Together Philosophy of Mathematics, Sociology of Mathematics, and Mathematics Education*, B. v. Kerkhove and J. P. v. Bendegem, Eds. Springer. Logic, Epistemology, and the Unity of Science, Vol. 5, 2007, pp. 135–151.

[2] ANDERSEN, P. B. Dynamic semiotics. *Semiotica 139*, 1/4 (2002), 161–210.

[3] ANDERSON, A. H., BADER, M., BARD, E. G., BOYLE, E., DOHERTY, G., GARROD, S., ISARD, S., KOWTKO, J., MCALLISTER, J., MILLER, J., ET AL. The HCRC Map Task Corpus. *Language and speech 34*, 4 (1991), 351–366.

[4] BARANY, M. '[B]ut this is blog maths and we're free to make up conventions as we go along': Polymath1 and the modalities of 'massively collaborative mathematics'. In *Proceedings of the 6th International Symposium on Wikis and Open Collaboration* (2010), ACM.

[5] BARWISE, J. Heterogeneous reasoning. In *Conceptual graphs for knowledge representation* (1993), Springer, pp. 64–74.


[16] http://folioharmonies.wordpress.com

[6] BARWISE, J., AND SELIGMAN, J. *Information flow: the logic of distributed systems*. Cambridge University Press, 1997.

[7] BENSON, B. E. *The improvisation of musical dialogue: a phenomenology of music*. Cambridge University Press, 2003.

[8] BORGO, D., AND GOGUEN, J. Rivers of consciousness: The nonlinear dynamics of free jazz. In *Jazz research proceedings yearbook* (2005), vol. 25.

[9] BUDZYNSKA, K. Circularity in ethotic structures. *Synthese 190*, 15 (2013), 3185–3207.

[10] BUDZYNSKA, K., AND REED, C. Speech acts of argumentation: Inference anchors and peripheral cues in dialogue. In *Computational Models of Natural Argument* (2011), F. Grasso, N. Green, and C. Reed, Eds. Papers from the 2011 AAAI Workshop.

[11] BUDZYNSKA, K., AND REED, C. The structure of ad hominem dialogues. In *COMMA* (2012), pp. 410–421.

[12] BUNDY, A. A science of reasoning: Extended Abstract. In *TABLEAUX '98 Proceedings of the International Conference on Automated Reasoning with Analytic Tableaux and Related Methods* (1998).

[13] BURKE, K., AND ZAPPEN, J. P. On persuasion, identification, and dialectical symmetry. *Philosophy and Rhetoric 39*, 4 (2006), 333–339.

[14] CANONNE, C., AND GARNIER, N. B. Cognition and segmentation in collective free improvisation: An exploratory study. In *Proceedings of the 12th International Conference for Music Perception and Cognition and 8th Conference of the European Society for the Cognitive Sciences of Music Joint Conference* (2012), E. Cambouropoulos, C. Tsougras, P. Mavromatis, and K. Pastiadis, Eds., pp. 197–204.

[15] CARLETTA, J., ISARD, A., ISARD, S., KOWTKO, J., DOHERTY-SNEDDON, G., AND ANDERSON, A. The coding of dialogue structure in a corpus. In *Proceedings of the Ninth Twente Workshop on Language Technology: Corpus-based Approaches to Dialogue Modelling. Universiteit Twente, Enschede* (1995).

[16] CARLSON, L. H. *Dialogue games: An approach to discourse analysis*. PhD thesis, Massachusetts Institute of Technology, 1982.

[17] COOK, N. Between process and product: Music and/as performance. *Music Theory Online 7*, 2 (2001), 1–31.

[18] CORNELI, J. *Peer produced peer learning: A mathematics case study*. PhD thesis, The Open University, 2014.

[19] DELEUZE, G. *Difference and repetition*. 1968. (trans. 2004, Paul Patton, London: Bloomsbury Academic).

[20] GANESALINGAM, M. *The Language of Mathematics*. Springer, 2013.

[21] GANESALINGAM, M., AND GOWERS, W. A fully automatic problem solver with human-style output. Tech. rep., 2013. arXiv:1309.4501v1.

[22] GOGUEN, J. An introduction to algebraic semiotics, with application to user interface design. In *Computation for metaphors, analogy, and agents*. Springer, 1999, pp. 242–291.

[23] GOGUEN, J. Musical qualia, context, time and emotion. *Journal of consciousness studies 11*, 3-4 (2004), 117–147.

[24] GOGUEN, J. Mathematical models of cognitive space and time. In *Reasoning and Cognition: Proc. of the Interdisciplinary Conference on Reasoning and Cognition* (2006), pp. 125–128.

[25] GORDON, M. From LCF to HOL: a short history. In *Proof, Language, and Interaction* (2000), pp. 169–186.

[26] GORDON, M. J., MILNER, R., AND WADSWORTH, C. *Edinburgh LCF: a mechanised logic of computation*, vol. 78 of *Lecture Notes in Computer Science*. Springer, 1979.

[27] GOWERS, W., AND GANESALINGAM, M. Modelling the mathematical discovery process, 2012. Maxwell Institute Lecture, Fri, November 2, 4pm – 5pm, James Clerk Maxwell Building, University of Edinburgh.

[28] GOWERS, W., AND NIELSEN, M. Massively collaborative mathematics. *Nature 461*, 7266 (2009), 879–881.

[29] HALL, E. T. Improvisation as an acquired, multilevel process. *Ethnomusicology* (1992), 223–235.

[30] HIRST, G. Context as a spurious concept. *arXiv preprint cmp-lg/9712003* (1997).

[31] HITCHCOCK, D. Informal logic and the concept of argument. *Philosophy of logic 5* (2006), 101–129.

[32] KAKAS, A., MAUDET, N., AND MORAITIS, P. Layered strategies and protocols for argumentation-based agent interaction. In *Argumentation in Multi-Agent Systems*. Springer, 2005, pp. 64–77.

[33] KILPATRICK, J. *Analyzing the Solution of Word Problems in Mathematics: An Exploratory Study*. PhD thesis, Stanford University, 1967.

[34] KOCKELMAN, P. Biosemiosis, technocognition, and sociogenesis: Selection and significance in a multiverse of sieving and serendipity. *Current Anthropology 52*, 5 (2011), 711–739.

[35] KOZA, J. R. *Genetic programming I: on the programming of computers by means of natural selection*. MIT press, 1992.

[36] KOZA, J. R. *Genetic programming II: automatic discovery of reusable programs*. MIT press, 1994.

[37] LAKATOS, I. *Proofs and refutations: The logic of mathematical discovery*. Cambridge university press, 1976.

[38] LARVOR, B. Albert lautman: Dialectics in mathematics. *Foundations of the Formal Sciences VII* (2011).

[39] LEECH, G., AND WEISSER, M. Generic speech act annotation for task-oriented dialogues. In *Procs. of the 2003 Corpus Linguistics Conference* (2003), Citeseer, Centre for Computer Corpus Research on Language, Lancaster University, pp. 441–446.

[40] LOUWERSE, M. M., AND CROSSLEY, S. A. Dialog act classification using n-gram algorithms. In *FLAIRS Conference* (2006), pp. 758–763.

[41] LUCAS, J. F. An exploratory study of the diagnostic teaching of heuristic problem solving strategies in calculus. In *Problem-Solving Studies in Mathematics*, J. G. Harvey and T. A. Romber, Eds., Wisconsin R&D Center Monograph Series. University of Wisconsin-Madison, School of Education, 1980.

[42] LUCAS, J. F., BRANCA, N., GOLDBERG, M., KANTOWSKI, M. G., KELLOG, H., AND SMITH, J. P. A process-sequence coding system for behavioral analysis of mathematical problem solving. In *Task Variables in Mathematical Problem Solving*, G. Goldin and E. McClintock, Eds. ERIC, 1980.

[43] MARTIN, J., AND GILLESPIE, A. A neo-meadian approach to human agency: Relating the social and the psychological in the ontogenesis of perspective-coordinating persons. *Integrative Psychological and Behavioral Science 44*, 3 (2010), 252–272.

[44] MEAD, G. H. *The philosophy of the present*. Open Court, 1932.

[45] MILNER, R. Logic for computable functions description of a machine implementation. Tech. rep., DTIC Document, 1972.

[46] MINSKY, M. *The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind*. Simon & Schuster, 2007.

[47] MINSKY, M. Essays on Education (for OLPC). Tech. rep., Massachusetts Institute of Technology, 2008–2009. Available online.

[48] MURRAY-RUST, D., AND ROBERTSON, D. Lscitter: building social machines by augmenting existing social networks with interaction models. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion* (2014), International World Wide Web Conferences Steering Committee, pp. 875–880.

[49] PARSONS, S., AND MCBURNEY, P. Argumentation-based communication between agents. In *Communication in Multiagent Systems*. Springer, 2003, pp. 164–178.

[50] PEASE, A., BUDZYNSKA, K., LAWRENCE, J., AND REED, C. Lakatos games for mathematical argument. In *Fifth International Conference on Computational Models of Argument* (2014), S. Parsons, N. Oren, and C. Reed, Eds., Frontiers in Artificial Intelligence and Applications, IOS Press.

[51] PEASE, A., AND MARTIN, U. Seventy four minutes of mathematics: An analysis of the third Mini-Polymath project. In *Proceedings of AISB/IACAP 2012, Symposium on Mathematical Practice and Cognition II* (2012).

[52] PELDSZUS, A., AND STEDE, M. From argument diagrams to argumentation mining in texts: a survey. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI) 7*, 1 (2013), 1–31.

[53] PÓLYA, G. *Mathematical discovery: On understanding, learning, and teaching problem solving*. John Wiley & Sons, 1981.

[54] PÓLYA, G. *Mathematics and plausible reasoning: Induction and analogy in mathematics*, vol. 1. Princeton University Press, 1990.

[55] PÓLYA, G. *Mathematics and Plausible Reasoning: Patterns of plausible inference*, vol. 2. Princeton University Press, 1990.

[56] PÓLYA, G. *How to solve it: A new aspect of mathematical method*. Princeton University Press, 2014.

[57] PRAKKEN, H. Formal systems for persuasion dialogue. *The Knowledge Engineering Review 21*, 02 (2006), 163–188.

[58] REED, C., AND BUDZYNSKA, K. How dialogues create arguments. In *Proceedings of the 7th Conference of the International Society for the Study of Argumentation (ISSA 2010)* (2011), F. van Eemeren et al., Eds., Sic Sat.

[59] REED, C., WELLS, S., BUDZYNSKA, K., AND DEVEREUX, J. Building arguments with argumentation: the role of illocutionary force in computational models of argument. In *COMMA* (2010), pp. 415–426.

[60] ROBERTSON, D. A lightweight coordination calculus for agent systems. In *Declarative agent languages and technologies II*. Springer, 2005, pp. 183–197.

[61] SCHOENFELD, A. H. Can heuristics be taught? *Cognitive process instruction* (1979), 315–338.

[62] SCHOENFELD, A. H. *Mathematical problem solving*. Academic Press, 1985.

[63] SCHOENFELD, A. H. What's all the fuss about metacognition. In *Cognitive science and mathematics education*, A. H. Schoenfeld, Ed. Lawrence Erlbaum Associates, 1987, pp. 189–215.

[64] SCHOENFELD, A. H. *How we think: A theory of goal-oriented decision making and its educational applications*. Routledge, 2010.

[65] SENNETT, R. *Together: the rituals, pleasures and politics of cooperation*. Yale University Press, 2012.

[66] SINGH, P. *EM-ONE: an architecture for reflective commonsense thinking*. PhD thesis, Massachusetts Institute of Technology, 2005.

[67] SINGH, P., AND MINSKY, M. An architecture for cognitive diversity. *Visions of mind: architectures for cognition and affect* (2005), 312.

[68] SOWA, J. F. Syntax, semantics, and pragmatics of contexts. In *Conceptual Structures: Applications, Implementation and Theory*. Springer, 1995, pp. 1–15.

[69] SPERBER, D., AND WILSON, D. Pragmatics, modularity and mind-reading. *Mind & Language 17*, 1-2 (2002), 3–23.

[70] STAB, C., AND GUREVYCH, I. Identifying argumentative discourse structures in persuasive essays. In *Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)* (Oct. 2014), Association for Computational Linguistics, p. (to appear).

[71] STAHL, G. *Group cognition: Computer support for building collaborative knowledge*. MIT Press Cambridge, MA, 2006.

[72] STAHL, G. *Studying Virtual Math Teams*, 1st edition ed. Springer, Sep 2010.

[73] STAHL, G. *Translating Euclid: Creating a human-centered mathematics*. Morgan & Claypool, 2013.

[74] STOLCKE, A., RIES, K., COCCARO, N., SHRIBERG, E., BATES, R., JURAFSKY, D., TAYLOR, P., MARTIN, R., VAN ESS-DYKEMA, C., AND METEER, M. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics 26*, 3 (2000), 339–373.

[75] THURSTON, W. On proof and progress in mathematics. *Bulletin (New Series) of the American Mathematical Society 30*, 2 (1994), 161–177.

[76] WALTON, D. How can logic best be applied to arguments? *Logic Journal of IGPL 5*, 4 (1997), 603–614.

[77] WALTON, D. Is there a burden of questioning? *Artificial Intelligence and Law 11*, 1 (2003), 1–43.

[78] WEBBER, B., EGG, M., AND KORDONI, V. Discourse structure and language technology. *Natural Language Engineering 18*, 4 (2012), 437–490.

[79] WINOGRAD, T. *Understanding natural language*. Academic Press, 1972.

[80] WOLSKA, M., AND KRUIJFF-KORBAYOVÁ, I. Analysis of mixed natural and symbolic language input in mathematical dialogs. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics* (2004), Association for Computational Linguistics, p. 25.

[81] YUAN, T., SCHULZE, J., DEVEREUX, J., AND REED, C. Towards an arguing agents competition: Building on argumento. In *Proceedings of IJCAI'2008 Workshop on Computational Models of Natural Argument* (2008).

# Appendix

**MiniPolymath1**:

Let $a_1, a_2, \ldots, a_n$ be distinct positive integers and let $M$ be a set of $n-1$ positive integers not containing $s = a_1 + a_2 + \ldots + a_n$. A grasshopper is to jump along the real axis, starting at the point 0 and making $n$ jumps to the right with lengths $a_1, a_2, \ldots, a_n$ in some order. Prove that the order can be chosen in such a way that the grasshopper never lands on any point in $M$.

**MiniPolymath2**:

In each of six boxes $B_1, B_2, B_3, B_4, B_5, B_6$ there is initially one coin. There are two types of operation allowed:

Type 1: Choose a nonempty box $B_j$ with $1 \leq j \leq 5$. Remove one coin from $B_j$ and add two coins to $B_{j+1}$.

Type 2: Choose a nonempty box $B_k$ with $1 \leq k \leq 4$. Remove one coin from $B_k$ and exchange the contents of (possibly empty) boxes $B_{k+1}$ and $B_{k+2}$.

Determine whether there is a finite sequence of such operations that results in boxes $B_1$, $B_2$, $B_3$, $B_4$, $B_5$ being empty and box $B_6$ containing exactly $2010^{2010^{2010}}$ coins. (Note that $a^{b^c} := a^{(b^c)}$.)

**MiniPolymath3**:

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 $\ell$ 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 $\ell$ going through $P$ such that the resulting windmill uses each point of $S$ as a pivot infinitely many times.

**MiniPolymath4**:

The liar's guessing game is a game played between two players $A$ and $B$. The rules of the game depend on two positive integers $k$ and $n$ which are known to both players.

At the start of the game, $A$ chooses two integers $x$ and $N$ with $1 \leq x \leq N$. Player $A$ keeps $x$ secret, and truthfully tells $N$ to player $B$. Player $B$ now tries to obtain information about $x$ by asking player A questions as follows. Each question consists of $B$ specifying an arbitrary set $S$ of positive integers (possibly one specified in a previous question), and asking $A$ whether $x$ belongs to $S$. Player $B$ may ask as many such questions as he wishes. After each question, player $A$ must immediately answer it with yes or no, but is allowed to lie as many times as she wishes; the only restriction is that, among any $k+1$ consecutive answers, at least one answer must be truthful.

After $B$ has asked as many questions as he wants, he must specify a set $X$ of at most $n$ positive integers. If $x$ belongs to $X$, then $B$ wins; otherwise, he loses. Prove that:

1. If $n \geq 2^k$, then $B$ can guarantee a win.
2. For all sufficiently large $k$, there exists an integer $n \geq 1.99^k$ such that $B$ cannot guarantee a win.