CrossMark

Argumentation Theory for Mathematical Argument

Joseph Corneli[1] · **Ursula Martin[1,2]** · **Dave Murray-Rust[3]** ·
Gabriela Rino Nesin[4] · **Alison Pease[5]**

Abstract

To adequately model mathematical arguments the analyst must be able to represent the mathematical objects under discussion and the relationships between them, as well as inferences drawn about these objects and relationships as the discourse unfolds. We introduce a framework with these properties, which has been used to analyse mathematical dialogues and expository texts. The framework can recover salient elements of discourse at, and within, the sentence level, as well as the way mathematical content connects to form larger argumentative structures. We show how the framework might be used to support computational reasoning, and argue that it provides a more natural way to examine the process of proving theorems than do Lamport's structured proofs.

Keywords Inference Anchoring Theory · Mathematical practice · Mathematical argument · Structured proof

1 Introduction

The representation of mathematical knowledge and inference in appropriate formal logical frameworks is well-understood and the subject of much research. Computational tools to support this through proof checking, automatic theorem proving, and

✉ Joseph Corneli
  joseph.corneli@ed.ac.uk

1   Present Address: School of Informatics, University of Edinburgh, Edinburgh, Scotland

2   Department of Computer Science, University of Oxford, Oxford, England

3   Edinburgh College of Art, University of Edinburgh, Edinburgh, Scotland

4   Computing, Engineering and Mathematics, University of Brighton, Brighton, England

5   Computing, School of Science and Engineering, University of Dundee, Dundee, Scotland

🍎 Springer

computer algebra are well-established, though they require formal, computationally explicit, content as input. However, the existing mathematical literature, particularly informal mathematical dialogues, and expository texts, is opaque to such systems, which cannot currently handle the variety of activities typically involved in producing such knowledge and proofs, such as, for example, exposition and argument that concerns making conjectures, forming concepts, and discussing examples and counterexamples. Our goal is to bridge this gap through devising an expressive modelling language that is closely related to the way mathematics is actually done.

Our approach to modelling such content is inspired by the general-purpose argument modelling formalism Inference Anchoring Theory (IAT), introduced by Reed and Budzynska (2010). As its name suggests, IAT anchors logical inferences in discourse. IAT has been applied to mediation (Janier and Reed 2017), debates (Budzynska et al. 2014b), and to paradoxes in ethotic argumentation (Budzynska 2013), along with other real-world dialogues (Budzynska et al. 2013). The *Inference Anchoring Theory + Content* (IATC) framework we introduce is based on IAT, but with several significant modifications. Most fundamentally, IATC is designed to bring to the surface the structural features inherent in mathematical content.

IATC could be overlaid upon formally specified contents, where these are available. Lamport's "Temporal Logic of Actions+" (TLA$^+$) (Lamport 1999, 2014) is one such formalism that could be used to model content-level expressions. Higher-level discourse structure would then be exhibited somewhat along the lines of Lamport's own semi-formal "structured proofs" (Lamport 1995, 2012). However, unlike structured proof, IATC does not aim to reshape the way people do mathematics, but to model it more exactly. As such, it constitutes groundwork for a future generation of computer systems that can collaborate with mathematicians and students in a way these potential users already understand. Epstein (2015) highlights the "extent to which a person believes that her work experience or product has been facilitated or improved by the collaboration" as a key evaluation metric for assessing collaborative intelligent computer systems. The key metric at this stage is more basic, namely, we are interested in the degree to which IATC can represent real-world examples of mathematical practice in a way that can make them accessible to computational reasoning. After introducing the modelling approach, we use several examples to show that IATC is indeed satisfactory in this regard.

- Our first example is a school-level challenge problem that was presented in a public lecture by the mathematician Timothy Gowers (Gowers and Ganesalingam 2012). The lecture aimed to motivate and contextualise a project, then beginning, to develop mathematical software that "operate[s] in a way that closely mirrors the way human mathematicians operate" (Ganesalingam and Gowers 2017, p. 255). The reasoning needed to solve the challenge problem remains beyond the scope of the computational method that Ganesalingam and Gowers ultimately published, but it is both sufficiently simple and sufficiently realistic to introduce the practical aspects of working with IATC.
- Our second example is a question posed on the online Q&A forum MathOverflow, together with the ensuing dialogue. MathOverflow is part of the Stack Exchange network of community question-and-answer websites, which is par-

ticularly popular with software developers. The MathOverflow sub-site is devoted to discussions about research-level questions in mathematics. Such discussions are very different from the textbook-style proofs treated by Ganesalingam and Gowers (2017), and we discuss the considerations that such discussions would impose on computational modelling efforts.

– MiniPolymath 1 through 4 were part of a series of experiments in collaborative online mathematics known as "Polymath projects" (Nielsen et al. 2009–2018). While other projects in the series tackled novel research, the problems in the MiniPolymath subseries were drawn from the Mathematical Olympiad, a premier competition for pre-college students. Six problems are given, and the examination takes place over two days with three problems to be solved each day. Whereas individual Olympiad participants frequently fail to solve three challenge problems in the four-and-a-half hours allotted for that purpose, all four of the collaborative MiniPolymath efforts generated a solution. However, it should be noted that some of these solutions took more than 24 hours to develop. IATC can help us understand how the proof efforts progressed, and can potentially help us understand why they were (mathematically) successful.

The plan of the work is as follows. Section 2 reviews previous research on mathematical argument, presents a brief introduction to Inference Anchoring Theory, and describes Lamport's structured proofs as an example of the state of the art for modelling informal mathematical knowledge. Section 3 introduces IATC, describes the grammar of IATC markup, and describes the differences between this language and IAT. Section 4 presents our analysis of the examples outlined above, which have been marked-up with IATC in order to illustrate the relevant modelling concerns. Section 5 summarises and reviews the contribution, situates our work in relationship to the broader literature, and outlines potential directions for further work.

2 Background

In this section we state what we mean by argumentation, and survey previous research on argumentation in mathematics (Sect. 2.1). We then describe Inference Anchoring Theory (Sect. 2.2) and structured proof (Sect. 2.3), two landmarks that guide our effort.

2.1 Argumentation and Mathematical Arguments

Our approach to argument builds on Buzynska and Reed's Inference Anchoring Theory (IAT), which we describe in Sect. 2.2. The specific conception of *argument* that underlies IAT is as follows:

> [A]rguing can be interpreted as an illocutionary act that comes about as the result of a relation between uttering a premise and uttering a conclusion, thus mirroring the logical structure of inference[.] (Reed et al. 2017, p. 146)

Reed and Budzynska (2010) note that in everyday language the term "argument" is used to describe a particular kind of interaction as well as the shared understanding extracted from these interactions, as "evidence" or "proof." The purpose of IAT is to make the links between discourse and reasoning explicit.

Concerning argumentation in a mathematical context, Pedemonte (2007, p. 39) argues that "analysis of the 'content' is not sufficient to analyse all the cognitive aspects in the relationship between argumentation and proof." A large part of mathematical discussion is in essence meta-discussion about meta-level objects, such as proof strategies that are suggested on the fly and debates about whether these strategies are likely to work as intended.

Mercier and Sperber (2011) distinguish arguments from inferences: only in the case of arguments "the reasons for drawing this conclusion on the basis of the premises are (at least partially) spelled out" (p. 58). By contrast, formal mathematics is typically based on the reductive assumption that "mathematical reasoning may be identified with classical, deductive inference" (Aliseda 2003, p. 25). However, everyday mathematical reasoning plainly involves more than just proof steps. Here are two examples of familiar patterns of reasoning that appear in MiniPolymath 3:

argument from authority	"My bachelor thesis supervisor said that one can't use the word cardinal if we talk about finite sets. One has to use the words 'number of elements'" (Tao et al. 2011, 19 July, 9:46 pm).
argument from analogy	"Let me check that I got the example correctly: is this 'a point inside a regular polygon'? Isn't it established in an early comment that the example of a point inside an equilateral triangle indeed visits *all* the points? Can you clarify the difference here?" (Tao et al. 2011, 19 July, 9:19 pm).

The word "argument" has been attached to several distinct kinds of mathematical artefacts and activities. This term may indicate *proofs* (Gasteren 1990), *informally-presented proofs* (Tanswell 2015), *proof sketches* (Lamport 1995), *aspects of reasoning that are not addressed by formal deduction* (Aberdein and Dove 2013) and *elements of persuasive discussion* (Zack and Graves 2001).

Some theorists have expressly contended that proofs are not arguments: this is because proofs offer certainty, while arguments cannot (Dufour 2013). Nevertheless, communication of reasons and reasoning can be found throughout mathematical practice. Pedemonte (2007) highlights the use of inductive and abductive logic as well as deduction in mathematical processes that move "from conjecturing to the construction of proof [to] the proof as product," and in which "content rather than formal criteria" can guide the proving process. Dufour (2013) gives examples of argumentation "not only before and during the proof but also after, at least as long as it can be criticized" (p. 74). Other scholars have observed features such as these:

- Published mathematical writing tends to be particularly explicit about reasons and conclusions (Dove 2009, p. 149).
- Not only the *Prover* but also the *Skeptic* "has an important role to play, namely to ensure that the proof is persuasive, perspicuous, and valid" (Dutilh Novaes 2016, p. 2618).
- On the way to a proof, degrees of confidence about the conclusions to be drawn may be discussed (Inglis et al. 2007, p. 17).
- Mathematical meanings need to be interpreted, and this tends to be a struggle (van Oers 2002, p. 360).

Carrascal (2015) provides an excellent survey of recent thinking about argument in mathematics, highlighting its connections with mathematical practice. Carrascal advises: "in order to learn more about the nature of mathematical practice and how its products are evaluated, we should be looking at real examples of this practice." She points to Pease and Martin (2012) as a notable example in this genre. Once we have developed a suitable apparatus, Sect. 4 will tackle several real-world examples, including a detailed reexamination of the dataset studied by Pease and Martin.

"Blog maths" (Barany 2010) and other online discussions, for example, on the question-and-answer site MathOverflow, can "tell us about mathematicians' attitudes to working together in public" as well as the "kinds of activities that go on in developing a proof" (Martin 2015). In the process of creating a proof or mathematical theory, divergent understandings are negotiated using shared concepts, definitions, and standards for proof, even as the concepts evolve. Along these lines, Pease et al. (2017) used the methods of structured and abstract argumentation to formalise the theory of informal mathematics developed in Lakatos's *Proofs and Refutations* (1976) as a set of rules for turn-taking in a dialogue game. This work shows that formally specified and fully implemented argumentation tools can be brought together and applied to a specific, demanding, domain of human reasoning.[1] Dauphin and Cramer (2018) produced a similar model of natural-deduction style arguments, explanations, and the "*prima facie* laws of logic" such as may be debated in work on mathematical foundations. These prior efforts focus on developing rules that give a plausible codification of mathematical process. Our concern is different, but complementary. We are interested in a better understanding of what is actually said in mathematical arguments, and on the reasoning that is conveyed. Accordingly, we will adapt a general-purpose argument modelling approach, *Inference Anchoring Theory*, which is described in the following section.

---

[1] The dialogue game defines ordered operations on a shared information state represented in the Argument Interchange Format (AIF) (Lawrence et al. 2012), which is then interpreted by The Online Argument Structures Tool (TOAST) (Snaith and Reed 2012) and passed on to DungOMatic (Snaith et al. 2010) to calculate the grounded extension, which in this case represents the currently accepted, collaboratively constructed, proof or theory under discussion.

Fig. 1 IAT diagram for the conversation '*A*'/'*Why?*'/'*A'*'
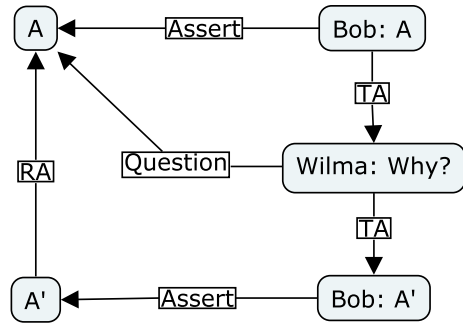
2.2 Inference Anchoring Theory

Inference Anchoring Theory (IAT) is used to model the logical relationships between the propositional contents of utterances made in dialogues (Budzynska and Reed 2011). As noted by Reed et al. (2017), the inspiration for developing IAT lies in earlier work on representing dialogue in the Argumentation Interchange Format.

IAT is grounded in a notion of *dialogical relations* that formalise the informal "conventions and norms that dictate the flow of dialogue" (Snaith and Reed 2016). Per Budzynska and Reed (2011), these dialogical relations are also referred to as "*transitions*," a term that is meant to recall the notion of transitions between operating states in a finite state machine. Indeed, when the norms have been fully codified in a *dialogue protocol*, the transitions are exactly described by a finite state machine.[2] Content relationships are typically identified by matching locutions against known argument schemes, e.g., an 'Argument from Positive Consequences' is associated with two transitions, 'challenging' and 'substantiating' (Walton et al. 2008). Budzynska et al. (2014a) describe Inference Anchoring Theory in terms of three components:

(i) relations between locutions in a dialogue, called *transitions*;
(ii) relations between sentences (propositional *contents* of locutions); and
(iii) *illocutionary connections* that link locutions with their contents.

(Budzynska et al. 2014a), emphasis added

In Figs. 1 and 2, below, "TA" stands for a *default transition*, "RA" stands for *application of rule of inference*, and "CA" stands for *default conflict*. That is to say, there is no explicit formal dialogue protocol attached to these two examples.

Figure 1 is a typical example of an IAT analysis. Figure 2 illustrates a feature that was not directly mentioned in the list (i)–(iii), above; specifically, this figure uses an *'implicit' speech-act* to anchor propositional content on a transition rather than a locution. Here, when a speaker asserts '*A*' and their interlocutor says '*No*', the

---

Fig. 2 IAT diagram for the conversation 'A'/'No'
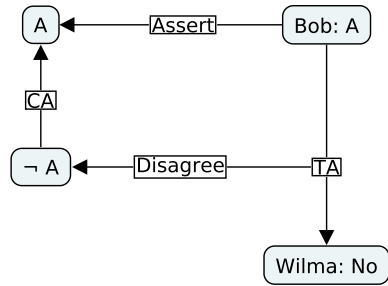
logical content '¬A' is attached to the transition, rather than to the negating word. The basic rationale is that the locution 'No' cannot be made sense of without the preceding context. There has been some debate about what to do about this. Botting (2015) says that the choice to anchor arguments on transitions is a conceptual mistake. However, for the creators of IAT, the reason illocutionary acts can be rooted on dialogical relations follows

> ...directly from pragma-dialectical analysis which views the speech act of assertion [...] as occurring at the 'sentence' level, and the speech act of argumentation as occurring at a 'higher textual level.' (Budzynska and Reed 2011)

Visser et al. (2011) describe the theoretical considerations in more detail. The pattern common to both Figs. 1 and 2 is that allowable inferences are governed by dialogue norms. In Fig. 1, for instance, we would not immediately know that 'A′' is intended to *support* 'A' without Wilma's intermediate question which explicitly requested such support. Given the context, the intended inference is clear. Thus, both examples serve to illustrate that

> the connection between locutions in a dialogue has an inferential component beyond any that may hold between the contents of those locutions (Reed and Budzynska 2010).

In short, IAT studies "the way in which the rules of dialogue influence the construction of argument" (Budzynska et al. 2016).

Although the specific example in Fig. 2 is very simple, the following general observation on dialogue norms is useful for thinking about how the conversation might continue from the point it has reached so far:

> [T]here is an asymmetry between the *production* of arguments, which involves an intrinsic bias in favor of the opinions or decisions of the arguer whether they are sound or not, and the *evaluation* of arguments, which aims at distinguishing good arguments from bad ones. (Mercier and Sperber 2011, p. 72)

If the conversation were to continue, Wilma would typically have the burden of justifying her rejection of 'A', which might be done with counterarguments that would dig into the details of 'A' looking for flaws (*ibid.*, p. 67); in addition, she might

begin to make a case for an alternative position, '*B*'. These considerations point to the direction we will be taking with IATC.

Our main strategy will be to supplement IAT with an explicit register for content. Alongside (i)–(iii), above, we introduce:

(iv) a model of non-propositional content, namely of the mathematical objects under discussion, and the relations between them.

We will describe the implications of this addition in detail in Sect. 3, along with some other adaptations to IAT that we have found useful in mathematical settings. One of the implications is that in the current work we do not need to emphasise transitions—of either the explicit or implicit variety—since a more explicit treatment of content gives us another way to manage context relationships.

2.3 Lamport's Structured Proofs

Structured proofs, as described by Lamport (1995, 2012), inhabit the middle ground between formal and informal mathematics, and provide a useful point of reference for our work on IATC. Structured proofs offer a notational strategy that is a "refinement of [...] natural deduction" (Lamport 1995). While the proofs represented using this system are not required to be strictly formal, the language of structured proofs has evolved together with Lamport's work on a formal language and corresponding proof checking system, the "Temporal Logic of Actions+" (TLA$^+$), which is used to model concurrent systems Lamport (1999, 2014).[3] Structured proofs are, specifically, structured as a strict hierarchy of lemmas. An example appears later on in this paper, in Fig. 6, which we will use to illustrate the similarities and differences with IATC.

For now, we comment that while the use of strict hierarchies is not representative of the way proofs are usually constructed in day-to-day practice, Lamport has proposed that structured proofs can assist in proof development, e.g., by helping to bring errors to the surface. However, they do not necessarily make the job of the reader easier: Lamport (2012, p. 20) quotes a referee who had read one of his structured proofs:

> The proofs [...] are lengthy, and are presented in a style which I find very tedious. [...] My feeling is that informal proof sketches [...] to explain the crucial ideas in each result would be more appropriate.

---

[3] Only a few of the keywords available in the latest version of TLA$^+$ appear in the structured proof notation. Per Lamport (2015), the full list of TLA$^+$ keywords is as follows. Those which are also used in structured proofs are decorated with underlining: <u>assume</u> ... <u>prove</u> ..., boolean, by, <u>case</u>, choose, constant (synonymously, constants), corollary, def, define, domain, else, except, extends, have, hide, if, instance, lambda, lemma, <u>let</u> ... <u>in</u> ..., <u>new</u>, omitted, <u>pick</u>, proposition, recursive, subset, <u>suffices</u>, take, theorem, unchanged, union, use, variable, witness.

Unlike structured proofs, IATC is intended to express the typical processes by which proofs are generated in standard practice, rather than make the process of proving and reading proofs easier. It would nevertheless be compatible with our aims to include formal statements in TLA$^+$ (or some other language) in IATC's content layer.

3 Inference Anchoring Theory + Content

IATC has many things in common with IAT, but should not be seen as a strict addition to the earlier theory. Adding explicit models of content and discussions about content prompts several adaptations. In this section we describe these adaptations, and introduce the IATC modelling language.

Several important requirements arise from the features of the mathematics domain. As we saw above, IAT is concerned with anchoring propositions to utterances and with mapping the logical relationships that obtain between them. However, various mathematical objects—Larvor (2012) mentions "diagrams, notational expressions, physical models, mental models and computer models"—are more comfortably thought of as non-propositional in nature. Discussions about proofs have been theorised formally using the notion of *proof plans*, which are constructed and transformed using explicit heuristics and tactics (Bundy 1988). However, Fiedler and Horacek (2007, pp. 63–64) have suggested that existing work with proof plans cannot be straightforwardly adapted from machine-oriented to human-oriented contexts, because proof plans are, from a potential human reader's perspective, overly detailed, with insufficient structural abstraction. By contrast, a language like IATC is charged with expressing "strategic arguments that are meaningful to humans" (Fiedler and Horacek 2007, p. 68). Nevertheless, as important as strategic reasoning is, low-level mathematical content seems to be even more fundamental.

We see the first-class role that content plays in mathematical discourse when new terms are introduced and referred to, for example. Thus, the editor's introduction to Karttunen (1976) notes the following:

> ...informal notational practise [sic] of mathematicians, who will write an existentially quantified formula (say, $(\exists e)(\forall x)(xe = ex = x)$, as one of a set of postulates for group theory) and thenceforth use the variable bound by the existential quantifier as if it were a constant as when they will write the next postulate $(\forall x)(\exists x^{-1})(xx^{-1} = x^{-1}x = e)$. [punctuation modified]

Karttunen's concept of "discourse referents," illustrated in the quote above, underlies Discourse Representation Theory (Kamp and Reyle 1993) and its extensions. While the developers of IAT acknowledge the generality of Structured Discourse Representation Theory (SDRT), in particular, they criticise it for making "assumptions of context-independent semantics" (Budzynska et al. 2016). Nevertheless, DRT has been successfully applied to model some aspects of mathematical discourse, and we will discuss that work further in Sect. 5, and contrast it with our orientation here.

For now, we emphasise that IATC differs from IAT in its approach to context. Specifically, IATC sets the notion of dialogical relations to one side, and instead
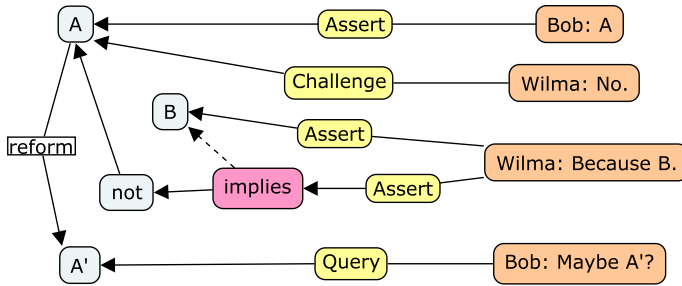
Fig. 3 Simple IATC diagram exhibiting an assertion, a refutation, a counterexample, and a reformation. (Color figure online)

connects locutions to each other directly in the content and intermediate (meta-discussion) layers.

Before we describe the language in detail, we present a simple example, Fig. 3, which reanalyses and extends the '*A*'/'*No*' dialogue from Fig. 2. The first two dialogue moves in these two examples are identical.

Here, rather than connecting '*No*' to '*A*' with a transition, we connect it directly to the previously modelled content, *A*, via a 'Challenge' illocution. From there, we continue to use the content and intermediate layers to explicitly model interconnections. For example, '*B*' does not simply conflict with *A*, but rather presents a warrant for "not *A*", modelled here using the two-parameter 'implies' relation.

With these changes in place, dialogue relations could in principle be reintroduced. For example, '*Because B*' could be seen to 'substantiate' the previous utterance, '*No*', as a communicated reason for rejecting *A*. Nevertheless, in the current work we continue to leave these links out, on the basis that we do not yet have a detailed theory of the norms of mathematical dialogue. The Lakatosian model developed by Pease et al. (2017), for example, only covers a limited subset of the rules and norms involved, specifically, those dealing with conjectures, lemmas, and the production and evaluation of counterexamples. By interconnecting contents in the content layer and through intermediate relations, we are able to make an explicit model of the logical structure of mathematical arguments. Such models could potentially inform a subsequent analysis of the associated dialogue structures.

For example, the long-range reform connection from *A* to *A*′ in our content analysis would suggest a corresponding long-range transition from Bob's first to his last statement in the dialogue. However, that would still neglect Bob's so-far implicit reasoning to the effect that *A*′ is (potentially) not vulnerable to objection *B*. If the dialogue continued from this point, detailed relationships between the constituent contents of '*A*′' and '*B*' may need to be discussed, and an IATC analysis would be able to unpack these and account for the details.

In line with these design decisions, and inspired by the specific features of mathematical dialogue and exposition, IATC introduces a range of extra machinery to the IAT framework to model the relationships between mathematical objects and propositions, along with an array of dialogue moves related to the strategic aspects of proof. Unlike IAT, we make no attempt to cover argumentation in law, natural

Table 1 Inference Anchoring Theory + Content, part 1: Performatives

Performatives (`perf[...]`)	
`Assert` (*s* [, *a*])	Assert belief that statement *s* is true, optionally because of *a*
`Agree` (*s* [, *a*])	Agree with a previous statement *s*, optionally because of *a*
`Challenge` (*s* [, *a*])	Assert belief that statement *s* is false, optionally because of *a*
`Retract` (*s* [, *a*])	Retract a previous statement *s*, optionally because of *a*
`Define` (*o*, *p*)	Define object *o* via property *p*
`Suggest` (*s*)	Suggest a strategy *s*
`Judge` (*s*, *v*)	Apply a heuristic value judgement *v* to some statement *s*
`Query` (*s*)	Ask for the truth value of statement *s*
`QueryE` ($\{p_i(X)\}$)	Ask for the class of objects *X* for which all the properties $\{p_i\}$ hold

science, or interpersonal mediation, fields in which the norms that govern inference can be vastly different. (Precedent, for example, may be acceptable in a legal argument but not in one about ethics.) In mathematical argumentation, many of the conventions are embodied in the objects under discussion and the things that can sensibly be said about them. Details of our notational apparatus are given in Tables 1 and 2. We collect reference examples of short texts marked up with these codes in an Appendix.
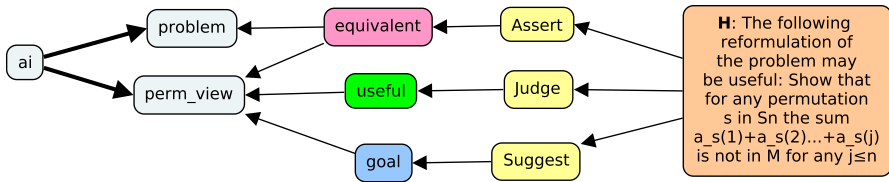
Our method for producing this set of tags was as follows. Two of us (with first degrees respectively in Mathematics and Information Systems, both with more than 10 years experience studying argumentation and social machines) performed close content analysis (Klaus 2004) together on the first 100 comments in MiniPolymath 1. Our analyses resulted in an initial tag set, including both typical illocutionary performatives and mathematics specific performatives, like `Define` and `QueryE`, as needed (see "Appendix" for examples). Several of the typical illocutionary connections (`Assert`, `Question`, `Challenge`, `Agree`) could be carried over from the schemes commonly applied in IAT. Our initial tag set was discussed and iteratively developed over the same 100 comments by all co-authors, with any recurring differences discussed, allowing us to align our results. A third co-author (with a first degree and Ph.D. in Mathematics) then further developed and refined the tag set by performing close content analysis on the entire MiniPolymath 3 conversation and on sections of MiniPolymath 1. Again, this was conducted alongside discussion with the other co-authors throughout the process. A fourth co-author (with a first degree in Mathematics) later extended the tag set with additional informal logical relationships, such as `analogy`, and specific content-focused relationships, such as `sums`, which played a role in the further examples we treated in Sect. 4. These extensions were again reviewed by all co-authors.

Our discussions concerned issues such as whether to label a statement such as '*it would be good to approach the problem in this way...*' as simply a suggested `strategy` or, additionally, as a `value[...]` judgement about the strategy. Shortly, in Fig. 4, we will show an example tagging in which the multiple layers of interpretation are included. However, perfect agreement about how to treat such cases is not intended; the IATC framework is designed to account

Table 2 Inference Anchoring Theory + Content, part 2: inferential structure, heuristics and value judgments, reasoning tactics, and content-focused relations

Inferential structure (`rel[...]`)	
`implies` (*s, t*)	Statement *s* implies statement *t*
`equivalent` (*s, t*)	Statement *s* implies statement *t* and vice versa
`not` (*s*)	Negation of *s*
`conjunction` (*s, t, ...*)	Conjunction of statements *s, t, ...*
`has_property` (*o, p*)	Object *o* has property *p*
`instance_of` (*o, m*)	Object *o* is an instance of the broader class *m*
`indep_of` (*o, d*)	Object *o* does not depend on the choice of object *d*
`case_split` (*s*, {*s_i*})	Statement *s* is equivalent to the conjunction of the *s_i*'s
`wlog` (*s, t*)	Statement *t* is equivalent to statement *s* but easier to prove
Heuristic value judgments (`value[...]`)	
`easy` (*s* [, *t*])	Statement *s* is easy to prove; optionally, easier than statement *t*
`plausible` (*s*)	Statement *s* is plausible
`beautiful` (*s*)	Statement *s* is beautiful (or mathematically elegant)
`useful` (*s*)	Statement *s* can be used in an eventual proof
Reasoning tactics (`meta[...]`)	
`goal` (*s*)	Used with `Suggest` to guide other agents to work to prove the statement *s*
`strategy` (*m, s*)	Indicate that method *m* might be used to prove *s*
`auxiliary` (*s, a*)	Statement *s* requires an auxiliary lemma *a*
`analogy` (*s, t*)	Statement *s* and statement *t* should be seen as analogous in some way
`implements` (*s, m*)	Statement *s* implements the method *m* from a previously suggested strategy
`generalise` (*m, n*)	Method *m* generalises method *n*
Content-focused structural relations (`struct[...]`)	
`used_in` (*o, s*)	Object *o* is used in statement (or object) *s*
`reform` (*s, t*)	Statement *s* can be reformed into statement *t*
`instantiates` (*s, t*)	Statement *s* schematically instantiates statement *t*
`expands` (*x, y*)	Expression *x* expands to expression *y*
`sums` (*x, y*)	Expression *x* sums to expression *y*
`cont_summand` (*x, y*)	Expression *x* contains *y* as a summand

for flexibility in interpretations. The additional tags in Table 2 were not at first divided into the present categories, but repeated analysis quickly revealed structural content relations, as well as inferential structure, as natural categories, intuitively corresponding to the mathematical and logical contents of the MiniPolymath discussions we examined. By far the most difficult categorisation to make was between value judgements and reasoning tactics. For example, the difference between deeming a statement `useful` and suggesting it as a `goal` could depend

```
perf[Assert](rel[equivalent](problem, perm_view))
perf[Judge](value[useful](perm_view))
perf[Suggest](meta[goal](perm_view))
struct[used_in](ai, problem)
struct[used_in](ai, perm_view)
```

Fig. 4 IATC markup of the statement "The following reformulation of the problem may be useful: show that for any permutation s in S_n, the sum $a_s(1) + a_s(2) \ldots + a_s(j)$ is not in M for any $j \leq n$." A larger portion of the dialogue is analysed in graphical and textual form in Fig. 12 and Table 3. (Color figure online)

completely on how polite or how bold the person making the utterance wished to be!

Our performatives have *slots*, which are filled by *statements* or *objects*. Statements may be represented in various ways: in unparsed natural language, as symbolic tokens that serve as shorthand for such statements, or in some representation language. The other relations are clustered into segments treating *Inferential Structure*, *Heuristics and Value Judgments*, *Reasoning Tactics*, and *Content-Focused Structural Relations*. The associated grammatical categories are given the following abbreviations in our linear notation: 'rel', 'value', 'meta', and struct'. For example, the expression 'perf[Assert](rel[has_property](o, p))' denotes the assertion of the statement "object o has property p." IATC allows direct, explicit, statements about objects, propositions, and statements. For example, 'perf[Assert](used_in (o, s))' denotes the assertion of the statement "object o appears in statement s."

We have two notational strategies that call attention to features of discourse or content that are taken as understood, but not explicitly stated. Performatives may be marked as "unspoken" when the contents are only broadly implied. Several examples of this notational strategy appear in Sect. 4.1. Similarly, content-focused structural relations are sometimes introduced without an attached performative, whenever they have been noticed by the analyst. Figure 4 includes examples of this latter usage. This figure represents the analysis of a short excerpt from a real mathematical dialogue, showing its diagrammatic and textual representations in IATC. The discussion ("MiniPolymath 1") concerned Problem 6 from the 2009 International Mathematical Olympiad. The text analysed in Fig. 4 is a portion of the fourth comment made in the discussion (Tao et al. 2009, 20 July, 6:50 am). An expanded excerpt is discussed in Sect. 4.3 along with more details of our IATC analysis of MiniPolymath data. Here, colour coding highlights the

correspondence between the graphical and textual grammar elements. One statement has been analysed into three performatives:

- The speaker `Asserts` that the problem has an *equivalent* reformulation. "The following **reformulation of the problem** may be useful**: Show that for any permutation** s **in** S_n**, the sum** $a_s(1) + a_s(2) \dots + a_s(j)$ **is not in** M **for any** $j \leq n$."
- The speaker `Judges` the reformulation to be (potentially) *useful*. "**The** following **reformulation** of the problem **may be useful**: [...]"
- The speaker `Suggests` that the reformulation describes a *goal* that could be worth pursuing: "[...] **Show that** [...]"

In addition, mathematical objects (several symbols, a_i) are analysed as component pieces of tagged content ('`problem`' and '`perm_view`'). Note that bold lines at left in the figure are a shorthand for the '`used_in`' relation. Subsequent statements in the dialogue will be able to link back to these objects: the analysis of an expanded extract appears in Fig. 12.

The relations given in Tables 1 and 2 have been sufficient to describe the reasoning in a range of examples, however we do not claim that this list of relationships would treat all mathematical texts. Nor do these relationships describe mathematical texts at the level of formality found in proof checking systems, or the level of detail found in some other theorisations of discourse. Thus, in the future IATC should not be limited to the set of tags presented here. For example, we have found uses for the value judgments '`easy`', '`beautiful`', and '`useful`', but it is quite plausible that future work would find use for values such as '`efficient`', '`generative`', or something else. Similarly, useful additions may be found in the other grammatical categories. The evidence from our examples in Sect. 4 is that these major grammatical categories—performatives, inferential relations, meta-level reasoning, value judgments, and content relations—are themselves stable.

We have described, and illustrated with simple examples, the way content and strategic relationships can be used to mediate contextual relationships, but context is also representable in IATC in another more explicit way. Although IATC does not require proofs to be structured in a tree-like hierarchy, nested structure is introduced as follows. In general, language elements in Table 2 that have a *statement* slot can also have that slot filled by a (possibly disconnected) *subgraph*. In this way, structure corresponding to a "lemma" can be indicated. A lemma, in this sense, is understood to be the reasoning that '`implements`' a '`strategy`', or, alternatively, a specific section of reasoning that '`implies`' some conclusion. This representation strategy is similar to the "partitioned networks" introduced by Hendrix (1975, 1979). An example will appear in Sect. 4.1.

To summarise, IATC resembles IAT in many ways, but with changes that are required when content, and discussions about content, are explicitly modelled. These features are necessary to express details of mathematical reasoning. For example, one proposition that can be extracted from the statement in Fig. 4 has the schematic form "The reformulation P is equivalent to the original question Q." IAT would have no way to extract P and Q from the assertion, but IATC can do so: they are represented as '`problem`' and '`perm_view`' in the figure. Later moves can

then connect to these pieces of content, and we already see such structure forming in our analysis of the above short excerpt.

IATC retains and extends IAT's approach to modelling contents and inferences, by adding non-propositional contents and more complex logical and heuristic relations. Illocutionary connections are also retained, with some mathematics-specific additions. However, IATC sets aside the notion of transitions, not because we view dialogue norms as unimportant, but because they are difficult to model at this stage. In IAT, relations between propositional contents roughly mirror the norms involved. The corresponding notion for IATC would be heuristics that account for the production of new expressions, and which take preceding expressions and background knowledge into account. We will have more to say about such heuristics in Sect. 4, nevertheless, many considerations must be deferred to future work.

4 Examples

In this section, we use three examples to showcase what IATC has to offer as a tool for analysis. We illustrate

– how IATC expresses the reasoning structures that arise in proof construction,
– how it might be used to support computational models of mathematical reasoning,
– and how it helps to uncover the salient elements of mathematical discourse.

To illustrate the points above, we have selected and analysed three examples that exhibit informal, expository, and discursive features of mathematical reasoning. The presentation here is a novel and self-contained synthesis and expansion of remarks made in previous papers (Corneli et al. 2017a, b; Pease and Martin 2012). The three examples collectively show the richness of mathematical argument, and were selected to match the three aims indicated above:

– Section 4.1: A carefully spelled out informal solution to a tricky but non-technical mathematical problem serves to illustrate the thought processes involved in successful mathematical problem solving. The example shows how IATC captures this sort of thinking.
– Section 4.2: A discussion of the relationships between, and merits of, different mathematical questions exhibits a level of abstraction above that needed in an individual proof. We explore the ramifications for explicit representations of the reasoning involved.
– Section 4.3: A multi-participant dialogue that develops a challenging but not highly technical proof casts light on processes of mathematical collaboration and mathematical reasoning. An analysis of this material using IATC allows us to explore the process of proof-construction in detail.

In each of the following subsections, we give more details of the context of each example, before presenting our analysis and comments.

What is the 500$^{\text{th}}$ digit of $(\sqrt{2}+\sqrt{3})^{2012}$?

Even this, eventually, a computer will be able to solve.

For now, notice that total stuckness can make you do desperate things. 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.

Can we do this for $(x+y)$? For e? Rationals with small denominator?

And how about small perturbations of these? Maybe it is close to a rational?

m^{th} digit of $(\sqrt{2}+\sqrt{3})^n$?

$(\sqrt{2}+\sqrt{3})^2$?

$(2+2\sqrt{2}\sqrt{3}+3)$

$(\sqrt{2}+\sqrt{3})^2 + (\sqrt{3}-\sqrt{2})^2 = 10$

$(\sqrt{2}+\sqrt{3})^{2012} + (\sqrt{3}-\sqrt{2})^{2012}$ is an integer!

And $(\sqrt{3}-\sqrt{2})^{2012}$ is a very small number. Maybe the final answer is "9"?

We need to check whether it's small enough. $(\sqrt{3}-\sqrt{2})^{2012} < \left(\frac{1}{2}\right)^{2012} = \left(\left(\frac{1}{2}\right)^4\right)^{503} = \left(\frac{1}{16}\right)^{503} < .1^{503}$, so we're in luck.

The answer is indeed 9.

Fig. 5 A "magic leap" challenge problem and its solution, presented by Timothy Gowers as part of a public lecture at the University of Edinburgh, November 2, 2012. (Reproduced from notes taken at the lecture.)

4.1 Making the Reasoning Explicit in the Solution to a Challenge Problem

In this section we aim to show that IATC is a natural modelling tool for informal mathematics. Whereas Robinson (1965, p. 23) had sought to

> reduce complex inferences, which are beyond the capacity of the human mind to grasp as single steps, to chains of simpler inferences, each of which is within the capacity of the human mind to grasp as a single transaction,

an alternative path of enquiry seeks to describe the heuristic process of proving theorems in more cognitively plausible terms. In particular, one relevant question to ask is how (human) mathematicians avoid large searches (Gowers 2017). IATC can contribute to the further development of this effort, by giving a uniform but expressive way to outline the process of developing proofs. Researchers working on mathematical software meant to exhibit human-style reasoning may find this expressiveness useful.

Our chosen example is a "magic leap" problem presented in a public lecture by Timothy Gowers, describing joint work with Mohan Ganesalingam (2012). The reasoning was communicated by a combination of speech and marks on a chalkboard, and is reproduced in Fig. 5. This example has been modelled in IATC by Corneli et al. (2017b). The problem initially appears difficult to solve without a computer algebra system, but a simple algebraic solution is available once the correct strategy is found. As such, an important part of the reasoning involved in solving the problem is to find the correct strategy. The steps involved in this part of the reasoning process are heuristic rather than deductive. We redescribe the analysis here.

PROOF SKETCH: What is the 500th digit of $(\sqrt{2}+\sqrt{3})^{2012}$? Even this, eventually, a computer will be able to solve. The fact that this has been set as a problem is a huge clue. Can we do this for $x+y$? For e? Small rationals? And how about small perturbations of these? Maybe it is close to a rational?

⟨1⟩1. For n large enough and m small enough in comparison, the mth digit   **1**
of a sufficiently small rational r to the nth power is equal to 0.
  PROOF:
  CASE: $r < 1/10$
  ⟨2⟩1. $(1/10)^n = .1^n$ has 0 in $n-1$ places in its decimal expansion.   **1.1**
  ⟨2⟩2. $r < 1/10$ implies r^n has zeros in at least $n-1$ places in its   **1.2**
    decimal expansion, so we simply need to select $m < n$. □

⟨1⟩2. Can we compute the mth digit of $(\sqrt{2}+\sqrt{3})^n$?   **2**
  PROOF:
  CASE: $n = 2$
  ⟨2⟩1. $(\sqrt{2}+\sqrt{3})^2 = 2 + 2\sqrt{2}\sqrt{3} + 3$ ∗   **2.1**

⟨1⟩3. Step 2.1 fails to give us an answer by direct computation, but if we   **3**
eliminate cross-terms, we can see that $(\sqrt{2}+\sqrt{3})^2$ is "close to" an
integer.
  PROOF:
  ⟨2⟩1. $(\sqrt{2}+\sqrt{3})^2 + (\sqrt{2}-\sqrt{3})^2 = 10$ **3.1**
  ⟨2⟩2. $(\sqrt{2}+\sqrt{3})^2 = 10 - (\sqrt{2}-\sqrt{3})^2$ **3.2**
  ⟨2⟩3. $(\sqrt{2}-\sqrt{3})^2$ is small   **3.3**
    PROOF:
    ⟨3⟩1. $(\sqrt{2}-\sqrt{3})^2 \approx (1\frac{1}{3}-1\frac{1}{2})^2$ by continued fraction approx-   **3.3.1**
      imation.
    ⟨3⟩2. $(1\frac{1}{3}-1\frac{1}{2})^2 = (\frac{8}{6}-\frac{9}{6})^2 = 1/36$ □   **3.3.2**

⟨1⟩4. Generalising Step 3.1, $(\sqrt{2}+\sqrt{3})^{2012} + (\sqrt{2}-\sqrt{3})^{2012}$ is an *integer*!   **4**
  PROOF:
  ⟨2⟩1. Terms with odd exponents cancel by the binomial theorem. □   **4.1**

⟨1⟩5. And $(\sqrt{2}-\sqrt{3})^{2012}$ is a *very* small number. Maybe the final answer   **5**
will be "9"? We need to check whether it's small enough.
  PROOF:
  ⟨2⟩1. $(\sqrt{2}-\sqrt{3})^{2012} < (1/2)^{2012}$ by Step 3.3.1   **5.1**
  ⟨2⟩2. $(1/2)^{2012} = ((1/2)^4)^{503}$ **5.2**
  ⟨2⟩3. $((1/2)^4)^{503} = (1/16)^{503}$ **5.3**
  ⟨2⟩4. $(\sqrt{2}-\sqrt{3})^{2012} < (1/16)^{503} < .1^{503}$, and so has at least 502   **5.4**
    zeros in its decimal expansion by Step 1
  ⟨2⟩5. The answer is indeed 9. □   **5.5**

Fig. 6 The solution to the challenge problem as a Lamport-style structured proof

For comparison with the IATC analysis, Fig. 6 reproduces the proof in Lamport's style. Figures 7, 8, 9 and 10 present portions of the IATC tagging of the solution that was presented in Gowers's lecture. Figure 7 illustrates an initial exploration of the question, and Fig. 8 establishes a 'strategy' based on that exploration ("The
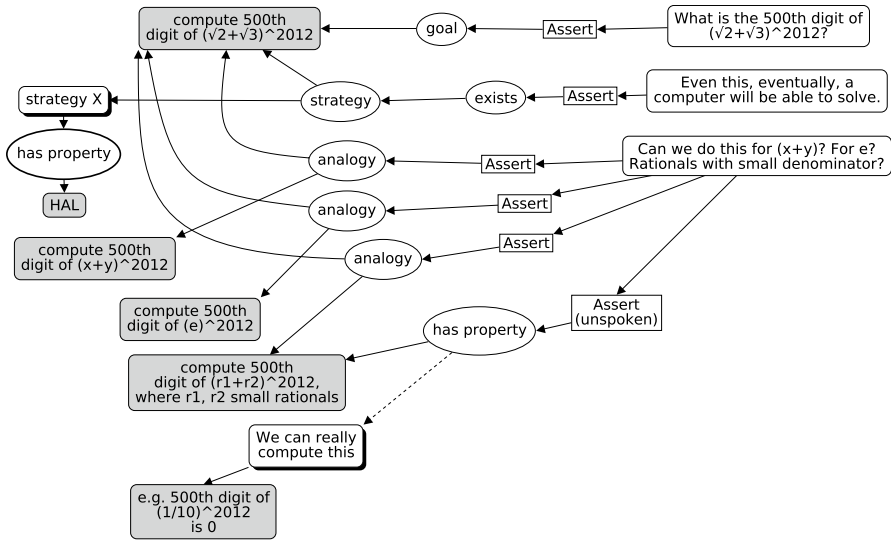
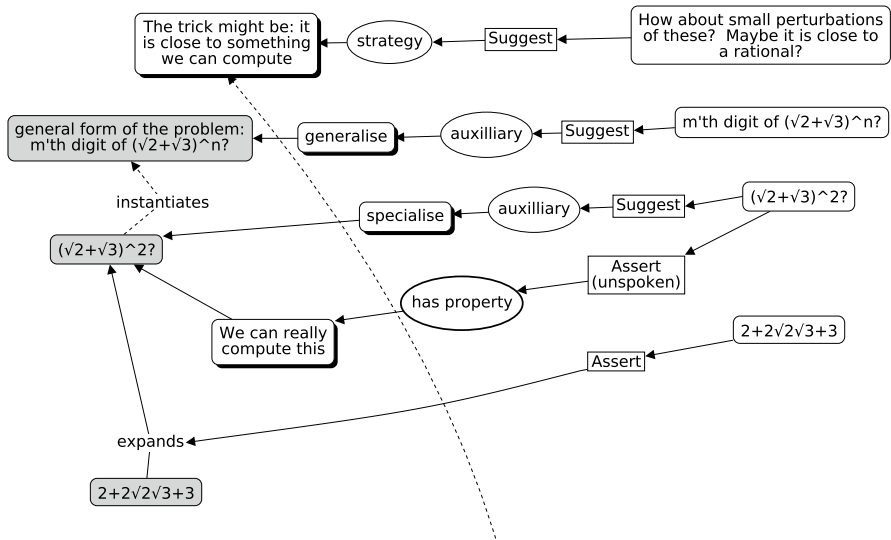**Fig. 7** IATC tagging for the first portion of the challenge problem

Fig. 8 IATC tagging for the second portion of the challenge problem

trick might be: it is close to something we can compute"). Figure 9 opens the door to applying the strategy. The central part of the proof that 'implements' the strategy is highlighted in Fig. 10.

The introduction to the proof, expanded in Fig. 7—and condensed into a "PROOF SKETCH" in Fig. 6—contains interesting examples of heuristic reasoning. This part of the solution centres on the probing question "Can we do this for 𝔛?", where 𝔛
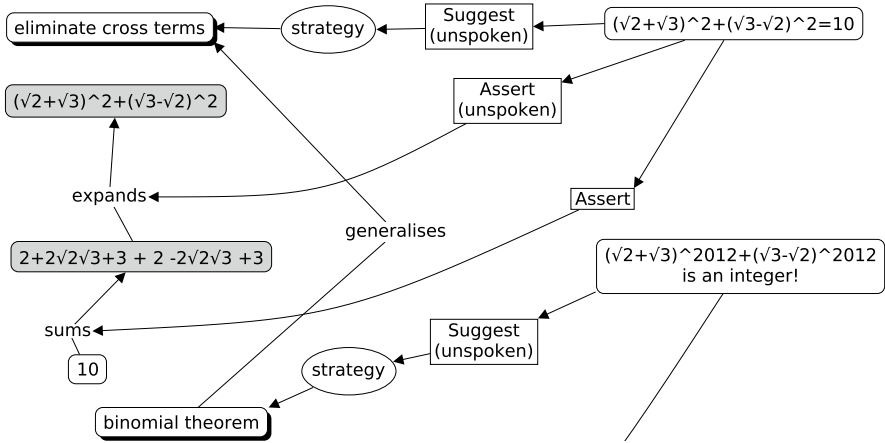
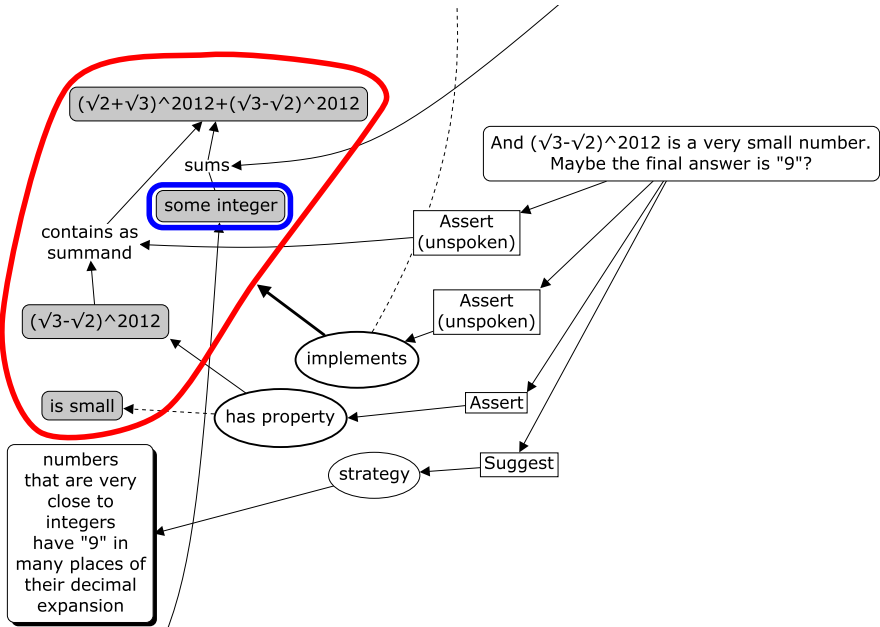**Fig. 9** IATC tagging for the third portion of the challenge problem

Fig. 10 Nested structure (in red) `implements` the `strategy` suggested earlier: "The trick might be: it is close to something we can compute." The intermediate conclusion reached in this phase of reasoning (highlighted in blue), when taken together with a further computational check, subsequently `implies` that the answer is "9". (Color figure online)

ranges over several examples: $x + y$, e, and small rationals, and where 'this' denotes "find the 500th digit of $\mathfrak{X}^{2012}$." In the IATC representation, each tentative proposal to "do this..." stands in `analogy` with the original problem statement. Although Fig. 7 contains only `Assert` performatives, a more complete representation would

also include `Query` performatives, since the analogies are not only proposed: their validity is also queried, much as we saw in the example treated in the previous section.

Step 1 in the structured proof works out one of the ideas from the proof sketch at a level of detail that was not present in the lecture, which instead progressed directly on to the material treated in Step 2. As Fiedler and Horacek (2007, p. 69) noted, "The analysis of human proof explanations shows that certain logical inferences are only conveyed implicitly, drawing on the discourse context and default expectations." There is no hard and fast rule that can tell us how much of the implicit material we need to explicate, but one rule of thumb that naturally arises from our representation strategy is that coherently related discussions should correspond to connected graphs in the expansion. Thus, for example, Fig. 7 includes an implicit "unspoken" `Assert`ion; the proof is made fully explicit in Step 1 of the Lamport-style proof, but never appeared in the original lecture. Again, in a standard IAT representation, unspoken assertions would typically be represented as 'implicit' speech acts rooted on transitions, whereas in IATC, we see how these unspoken assertions play a role in the argument via their expansion and subsequent interconnections in the content layer.

Indeed, nowhere in the explicitly communicated reasoning is the key `strategy` fully and explicitly stated. The basic strategy of the proof is that *the quantity of interest may be sufficiently close to something we can compute*. In the IATC representation (Fig. 8), this is understood to be `Suggest`ed by the following statements from the proof sketch, "And how about small perturbations of these? Maybe it is close to a rational?" Step 1 of the structured proof shows that rationals do, in fact, match the strategy's preconditions. The IATC representation is less explicit on this point, since it sticks more closely to the reasoning expressed in the lecture. This example shows that even relatively explicit statements may need further interpretation to be represented meaningfully in IATC. Specifically, the way the proof progresses only makes sense if we recognise the '`strategy`' implied by what might otherwise appear to be a throwaway comment early on.

Step 2 in the structured proof concerns another analogy. This time, a special one which, the IATC analysis notes, symbolically `generalise`s the initial question (Fig. 8). That is, rather than considering $(\sqrt{2} + \sqrt{3})^{2012}$ we now consider $(\sqrt{2} + \sqrt{3})^m$. (NB. an edge connecting the '`generalise`' node to the problem statement has been omitted.) However, the concept of generalisation remains implicit in the corresponding portion of the structured proof. Indeed, Step 2 is not a good match for the requirements of structured proof at all, since it is not a real lemma, and its "proof" fails (indicated by "*"). Including failed proof steps is not a problem for IATC. In Fig. 9 the process of solving the problem proceeds apace, without pausing to remark on a failed lemma, now that something more interesting has been discovered.

Meanwhile, Step 3 in the structured proof implements the main strategy for resolving a special case of our generalised problem, namely showing that $(\sqrt{2} + \sqrt{3})^2$ is close to an integer, establishing a pattern that leads to the conclusion. Again, Step 3.3 offers considerably more detail than was present in the original lecture.

Step 4 subsequently generalises the method that was used in Step 3, and applies it to the expression we were originally interested in. Figure 10 diagrams out the reasoning that underlies this step. The long-range dashed edge in this figure connects with the node "The trick might be: it is close to something we can compute" pictured in Fig. 8. The collection of nodes highlighted in red `imple-ment` that strategy. Notice, though, that the computation is not done explicitly: it's unimportant which integer the number of interest is close to. Collectively, the fact that $(\sqrt{2} + \sqrt{3})^{2012} + (\sqrt{3} - \sqrt{2})^{2012}$ sums to "some integer" and the fact that $(\sqrt{3} - \sqrt{2})^{2012}$ is sufficiently small `implies` the result. Step 5 shows the details of the final computational check.

Several objections could be raised about the structured proof presented in Fig. 6, most notably to the inclusion of a failed lemma in Step 2. However, as a source of information about the intuition behind the proof, this failure is valuable. While objections to the IATC treatment are also possible, it is clear that this method helps to make explicit features of the proof process that remain implicit in the structured proof. In particular, analogies, strategies, and relationships between methods are made explicit. While the structured proof augments the lecture with more technical details, IATC provides a more faithful model of the reasoning expressed in the lecture itself.

4.2 Towards Computable Models of Mathematical Reasoning Via IATC: A Q&A Example

Contributors to discussions about mathematics on MathOverflow do more than just talk about proofs.

> The presentation is often speculative and informal, a style which would have no place in a research paper, reinforced by conversational devices that are accepting of error and invite challenge. (Martin and Pease 2013)

IATC allows the argumentation aspects of mathematical dialogues to be represented as explicit graphical structures, which gives a plausible basis from which to develop an explicit computational model of the reasoning steps that are implied in mathematical argumentation. Corneli et al. (2017a) showed how IATC could be used to create graphical models of the discussion that develops around a question posted on MathOverflow. Here we will remark further on implications for computational modelling. The question, which was given the title "Group cannot be the union of conjugates" (Chandrasekhar et al. 2010), is as follows:

"*I have seen this problem, that if G is a finite group and H is a proper subgroup of G with finite index then $G \neq \bigcup_{g \in G} gHg^{-1}$. Does this remain true for the infinite case also?*"

In the most straightforward reading, two superficially similar group-theoretic propositions seem to be at stake:

(*P1*)   "*If G is a <u>finite</u> group, H is a subgroup of G and the index $[G : H]$ is finite, then G is not equal to the union of gHg^{-1}*"; and,

(P2)   "*If G is an <u>infinite</u> group, H is a subgroup of G and the index $[G : H]$ is finite, then G is not equal to the union of gHg^{-1}.*"

The question thus implicitly outlines an argument by analogy:

– (*P1*) is true
– (*P2*) is similar to (*P1*)
– Therefore, (*P2*) is (potentially) true as well

The essence of the question is to ask whether the mathematical facts align with this schematic argument. As it turns out, this question is answered in the affirmative. Shortly after the question was asked, one discussant make the terse comment "the case of infinite *G* readily reduces to the case of finite *G*"; months later, another discussant supplies an explicit proof of (*P2*).

In the mean time, other discussants had proposed and addressed several alternative formulations of the question. An important distinction hinges on the interpretation of the phrase "*infinite case.*" An alternative proposition that incorporates some of the suggested revisions is as follows:

(P2′)   "*If G is an <u>infinite</u> group, H is a proper finite index subset of G and the index $[G : H]$ is <u>infinite</u>, then G is not equal to the union of gHg^{-1}.*"

In this case an argument by analogy would not match the facts: a counterexample is supplied to show that proposition (*P2′*) is false.

The dialogue is an interesting example of mathematical reasoning in which proof certainly plays a role, but is nevertheless of secondary interest compared with asking interesting questions, and thinking about how different questions relate to each other. What would be necessary to represent this sort of dialogue computationally? Expressing propositions like (*P1*) in IATC is straightforward, though, as we noted, the content layer is not directly modelled in this representation language. The following expression represents this proposition in IATC, introducing additional invented pseudocode representations (in italics) in the content layer.

```
perf[Assert](
 rel[implies](
  rel[conjunction](finite_group(G),
                   subgroup(H,G),
                   rel[has_property](index(H,G),is_finite))
  rel[not](equal(G,
                 union_over(conjugates(H,g),elements(g,G)))))))
```

Processing such expressions to build a model of a dialogue will require adding numerous stanzas like this one, each rooted on an IATC performative, into one graph database that records the relationships between the statements and their constituent parts. Individual expressions like the `implies` relationship would need to be addressable, in order for an `analogy` between two implications to be

proposed. Definitions for predicates like `finite_group` and special constructions like `union_over` could be supplied in an accompanying knowledge base. In further rounds of computational processing, the analogies between (*P1*) and (*P2*), and between (*P1*) and (*P2′*), could be checked using graph-processing methods described by Sowa and Majumdar (2003). New heuristics would be needed if the aim was to demonstrate the truth or falsity of the various propositions, not just to recreate the surface analogies. Moreover, as we've seen, mathematical dialogues are not just concerned with verifying statements, but may also consider the qualities that make a particular question interesting in a given context. Heuristics that can be used to select interesting problems are not prevalent in current mathematical software.

As a limited proof of concept showing the plausibility of adding a computational deduction and verification layer on top of IATC representations, Corneli et al. (2017b) give a detailed expansion of one step of a mathematical proof using simple rules for transforming the underlying graph structures. It is worth emphasising that the representations of reasoning afforded by language elements in Tables 1 and 2 do not themselves encode the meta-level reasoning associated with such graph transformations.

4.3 MiniPolymath Revisited

The data that underlie this section were generated in a series of online experiments in collaborative problem solving convened by mathematician Terence Tao (2009; 2011). We use IATC to expand on a previous analysis of this data presented by Pease and Martin (2012), showing how IATC can advance the theory of mathematical argument through the detailed analysis of real world examples, as per Carrascal (2015).

In their 2012 paper, Pease and Martin analysed the third MiniPolymath project in broad strokes, with each blog comment comprising a single unit to be tagged. They developed a typology of five intuitive comment types, based on the mathematical content of each comment: *examples*, *conjectures*, *concepts*, *proofs*, and *other*.

In order to assign comments to these categories, both authors performed close content analysis on all comments posted between the time Tao posted the problem to his blog (8pm, UTC on July 19th, 2011) and the time he announced that a solution had appeared (9.50pm, UTC on July 19th, 2011). The discussion comprised 147 comments over 27 threads. Ten comments were assigned to more than one category.

Our present IATC analysis of the same data is designed to give a more complete picture of the linguistic, dialectical, and inferential structure of the comments that fall within the five intuitive categories mentioned. There are three main differences between the two analyses. First, in comparison with the earlier broad-stroke analysis, the IATC analysis is richly detailed, with a unit defined as any quantum of commentary with taggable content. Secondly, our focus in the earlier analysis was purely on mathematical content, and on the *type* of mathematical content in particular. This contrasts with our present analysis, in which we provide a more fine-grained representation of mathematical content in the taggable units, and furthermore take into account linguistic, dialectical, and inferential structure. Third, the IATC analysis
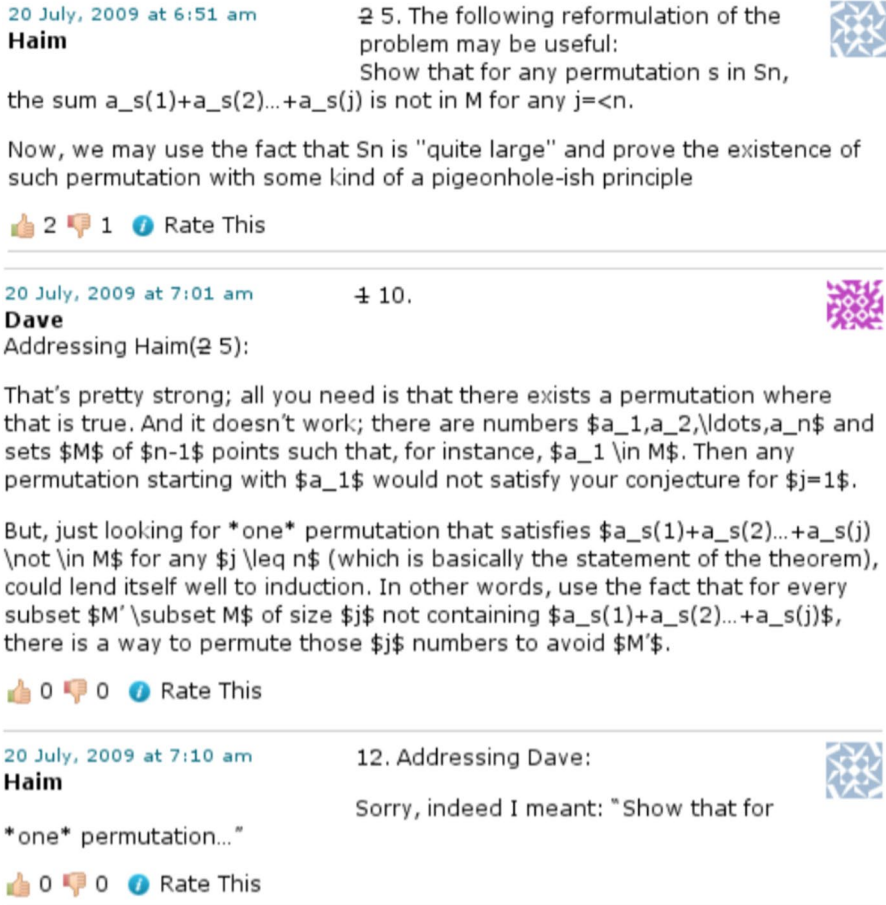
20 July, 2009 at 6:51 am
Haim

2̶ 5. The following reformulation of the problem may be useful:
Show that for any permutation s in Sn,
the sum a_s(1)+a_s(2)...+a_s(j) is not in M for any j=<n.

Now, we may use the fact that Sn is "quite large" and prove the existence of such permutation with some kind of a pigeonhole-ish principle

👍 2 👎 1 ⓘ Rate This

---

20 July, 2009 at 7:01 am
Dave
Addressing Haim(2̶ 5):

± 10.

That's pretty strong; all you need is that there exists a permutation where that is true. And it doesn't work; there are numbers $a_1,a_2,\ldots,a_n$ and sets M of $n-1$ points such that, for instance, $a_1 \in M$. Then any permutation starting with a_1 would not satisfy your conjecture for $j=1$.

But, just looking for *one* permutation that satisfies $a_s(1)+a_s(2)...+a_s(j) \not \in M$ for any $j \leq n$ (which is basically the statement of the theorem), could lend itself well to induction. In other words, use the fact that for every subset $M' \subset M$ of size j not containing $a_s(1)+a_s(2)...+a_s(j)$, there is a way to permute those j numbers to avoid M'.

👍 0 👎 0 ⓘ Rate This

---

20 July, 2009 at 7:10 am
Haim

12. Addressing Dave:

Sorry, indeed I meant: "Show that for *one* permutation..."

👍 0 👎 0 ⓘ Rate This

Fig. 11 Screenshot of a portion of the MiniPolymath 1 dialogue. (Color figure online)

takes into consideration the entire MiniPolymath 3 conversation, including the comments that came after Tao had announced that a proof had been found.

The new analysis, accordingly, adds depth to our earlier analysis. Crucially, the new perspective will be more relevant to argumentation theorists, and supports a detailed understanding of what went on in the process of constructing the collaborative proof. The earlier typology provided an initial way to sort the content, whereas the IATC tag set developed along with our analysis via the iterative, discursive method discribed in Sect. 3. Though they cover the same data and show some correlations, as described below, the latter categorisation was not derived from the earlier one.

Figure 11 presents an excerpt from the MiniPolymath 1 dialogue (MPM1) as it originally appeared on Tao's blog. Figure 12 and Table 3 give the IATC analysis of this excerpt in diagrammatic and textual form. The first portion of Fig. 12
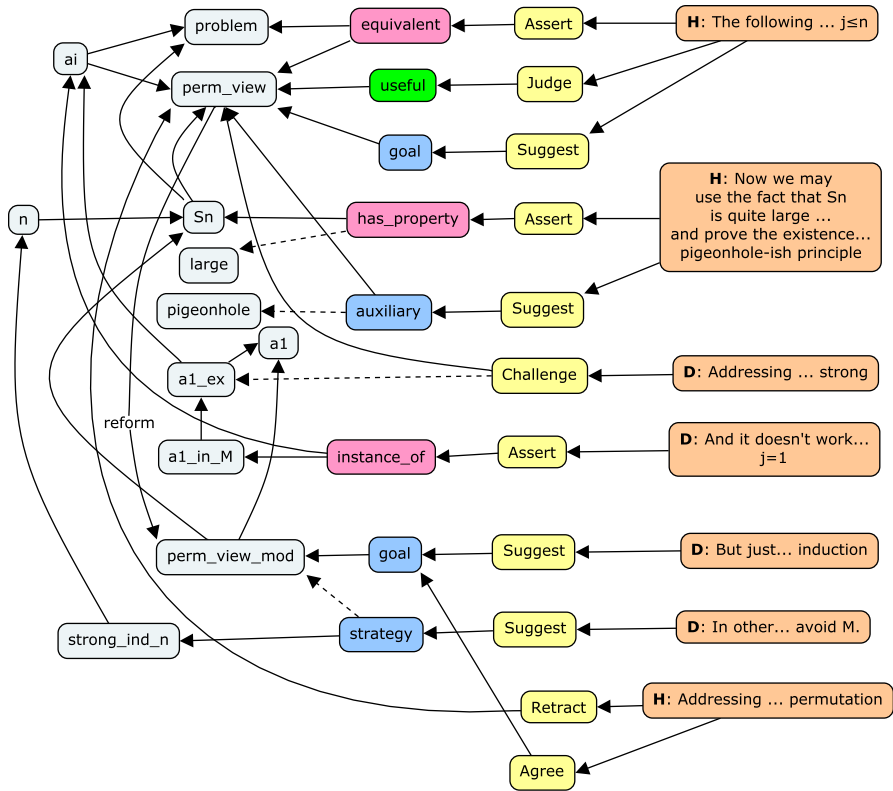
Fig. 12 IATC analysis of MPM1 excerpt (graphical form). (Color figure online)

repeats the contents of Fig. 4. The longer excerpt shown here illustrates complex contextual interconnections forming in the content layer.

Our main example in this section is MiniPolymath 3 (MPM3), which we tagged into IATC in its entirety. (This work was carried out by one co-author with a first degree and PhD in Mathematics, in consultation with others as described in Sect. 3.) As an indicative sample, the first three comments and their tags are shown in Fig. 13. Figure 14 shows how tags from IATC's five grammatical categories were distributed over time. Thus, for example, we see 'value' tags used early in the discussion as strategies are being considered, and again later in the discussion when solutions are being vetted. Figure 15 gives another view of the timeline, showing how the comments were categorised into the 5-part typology from Pease and Martin. In the initial categorisation developed for that paper, comments were allowed to be in multiple categories at once. Here, to facilitate a clean mapping to IATC, we redid the categorisation with the requirement that each comment should fit into exactly one main category. We arrived at a nearly equal division of comments among the five categories: *example* (20.3%) *conjecture* (21.2%), *concept* (19.5%), *proof* (19.5%), and *other* (19.5%). (This

Table 3 IATC analysis of MPM1 excerpt (text form). (Color table online)

(**Haim**) The following reformulation of the problem may be useful: "Show that for any permutation s in S_n, the sum $a_{s(1)} + a_{s(2)} + \ldots + a_{s(j)}$ is not in M for any $j \leq n$."

(**1**) `perf[Assert](rel[equivalent](problem, perm_view)),`
(**2**) `perf[Judge](value[useful](perm_view)),`
(**3**) `perf[Suggest](meta[goal](perm_view)),`
(**4**) `struct[used_in](ai, problem),`
(**5**) `struct[used_in](ai, perm_view)`

(——) Now, we may use the fact that S_n is "quite large" and prove the existence of such permutation

(**6**) `perf[Assert](rel[has_property](Sn, large)),`
(**7**) `struct[used_in](Sn, perm_view),`
(**8**) `struct[used_in](Sn, problem)`

(——) with some kind of a pigeonhole-ish principle.

(**9**) `perf[Suggest](meta[auxiliary](meta[strategy]( perm_view,pigeonhole), rel[has_property](Sn, large)))`

(**Dave**) Addressing Haim(2 5): That's pretty strong. And it doesn't work; there are numbers $a_1, a_2, \ldots, a_n$ and sets M of $n-1$ points such that, for instance, $a_1 \in M$. Then any permutation starting with a_1 would not satisfy your conjecture for $j = 1$.

(**10**) `perf[Challenge](perm_view, a1_ex),`
(**11**) `perf[Assert](rel[instance_of](a1, a1_in_M)),`
(**12**) `struct[used_in](a1_ex, a1_in_M)`

(——) But, just looking for **one** permutation that satisfies $a_{s(1)} + a_{s(2)} + \ldots + a_{s(j)} \notin M$ for any $j \leq n$ (which is basically the statement of the theorem), could lend itself well to induction.

(**13**) `perf[Suggest](meta[goal](perm_view_mod)),`
(**14**) `struct[reform](perm_view, perm_view_mod),`
(**15**) `struct[used_in](Sn, perm_view_mod),`
(**16**) `struct[used_in](a1, perm_view_mod)`

(——) In other words, use the fact that for every subset $M' \subset M$ of size j not containing $a_{s(1)} + a_{s(2)} + \ldots + a_{s(j)}$, there is a way to permute those j numbers to avoid M'.

(**17**) `perf[Suggest](meta[strategy](perm_view_mod, strong_ind_n)),`
(**18**) `struct[used_in](n, strong_ind_n),`
(**19**) `struct[used_in](n, Sn)`

(**Haim**) Addressing Dave: Sorry, indeed I meant: "Show that for **one** permutation..."

(**20**) `perf[Retract](perm_view),`
(**21**) `perf[Agree](meta[goal](perm_view_mod))`

"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 ∈ 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."

```
rel[structural](problem, S)
rel[structural](problem, P)
rel[structural](problem, l)
rel[structural](problem, windmill)
rel[structural](problem, pivot)
rel[structural](P, pivot)
rel[structural](P, windmill)
rel[structural](l, windmill)
```

1. *Could you start off with a random point in the plane and prove it doesn't work, if you can't prove that then the opposite holds.*

```
perf[query](random_test_false)
perf[assert](rel[stronger](rel[not](prove_rtf),
                           rel[not](random_test_false)))
rel[structural](random_test_false, prove_rtf)
```

2. *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.*

```
perf[define](pivot_seq, ps_def)
rel[structural](pivot_seq, pivot)
```

2.1. *Nice. We need only to consider the times when to points are connected – this gives us a path, and after some time this path will come back to some already visited point. So there is a cycle. If only we could find a cycle which spans all the points, the question is solved. . . That may be some useful simplification.*

```
perf[assert](rel[has_property](pivot_seq, has_cycle))
perf[suggest](meta[goal](cycle_spans_S))
perf[judge](value[useful](pivot_seq))
rel[structural](S, cycle_spans_S)
```

Fig. 13 IATC tags for the problem and first three comments in MiniPolymath 3

replication work was carried out independently by one of the coauthors with a first degree in Mathematics.)

Figure 16 illustrates the correspondence between IATC tags with the earlier typology. Aligning the bulkier 5-part categorisation with the IATC tagging shows
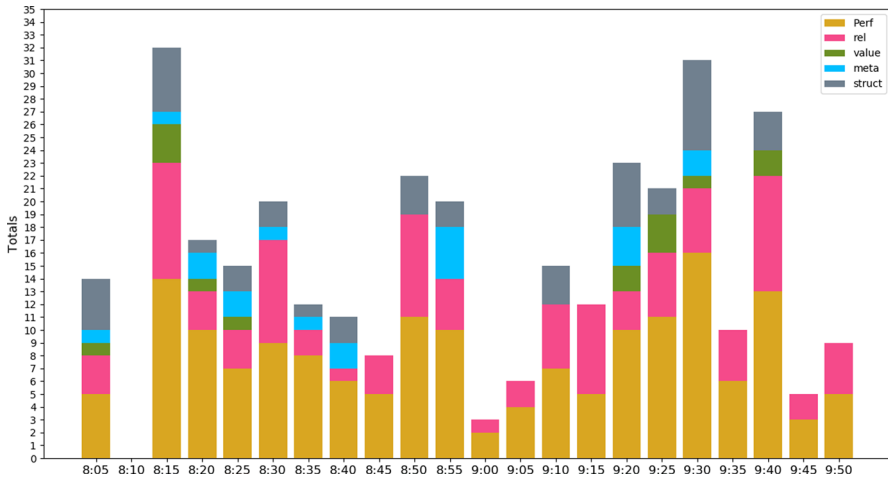
Fig. 14 Timeline of the MiniPolymath 3 dialogue showing the IATC grammar categories used in the tagging. Comments are binned into 5 min intervals. The first interval is 8:05–8:09 and the last is 9:50–9:59, inclusive. (Color figure online)
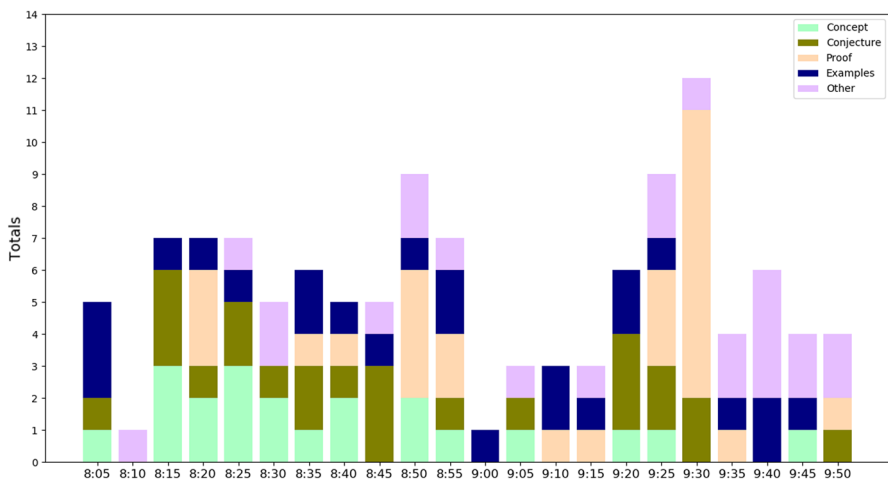

Fig. 15 Timeline of the MiniPolymath 3 dialogue showing comments categorised into five categories: Concept, Conjecture, Proof, Examples, and Other. (Color figure online)

that these five intuitive labels are mapped in very different ways to the more detailed IATC tag set.

We observe certain regularities: for example, `Assert` is present in all five types of comments, but is used most frequently within proof-related comments. Annotations from the 'struct' grammatical category are most prevalently associated with conjecture-related comments. (NB. In this tagging exercise we only considered the 'used_in' facet of the 'struct' category, so 'structural'

is here a synonym for 'used_in'.) It is not surprising that the performative `Challenge` is used most frequently in examples, since, intuitively, an example is likely to be put forward as a counter-example. The most prevalent use of `Agree` is in comments that are categorised as "other". `Retract` is frequently used in this category as well, as is `stronger` (here, a synonym for 'implies'). These usages reflect social values as well as mathematical semantics. E.g., one can express support for an idea by underscoring one's belief in an implication, as in the comment "Yes, it seems to be a correct solution!" (Tao et al. 2011, July 19, 9:35 pm).

One might suspect that `Suggest` should be used only within conjectures, but in the current categorisation it is used somewhat more frequently along with concepts. This is partly explained by the fact that `Suggest` can be used to introduce either a `goal` or a `strategy`. Sometimes goals represent conceptual tidying, as in "I guess there is an odd/even number of point distinction to do" (Tao et al. 2011, July 19, 9:31 pm).

Furthermore, despite our self-imposed constraint to map each comment only to the most salient of the five categories, in practice a comment may simultaneously introduce a concept along with a conjecture that applies that concept. For example the straightforward concept of "restriction[s] on how the next pivot is chosen" appears along with the more speculative conjecture "Can we start with a complete graph and all cycles on that graph and just discard the ones that don't follow the restrictions to converge on the ones that do?" (Tao et al. 2011, July 19, 8:56 pm). The need to introduce concepts also applies in the case of more outlandish conjectures, such as "It might be fun to use projective duality" (Tao et al. 2011, July 19, 8:23 pm). However, a concept may suggest a vague method without raising a conjecture as such, e.g., "I'm thinking spirograph rather than convex hull" (Tao et al. 2011, July 19, 8:44 pm).

In sum, the IATC analysis of MiniPolymath 3 shows in detail how individual contributions to the dialogue are comprised. In aggregate, this analysis exposes the structural anatomy of a successful collaborative proof. It should be noted that not all the contributions to MPM3 were equally relevant to the final solution. By entering the structures in an explicit graphical model in the manner described in Sect. 4.2, graph theoretic analysis could establish, e.g., the centrality of the various concepts used in the content layer, and who introduced them into the conversation.

5 Conclusion

We have sought to advance the study of mathematical practice from an argumentation-theoretic perspective. We introduced Inference Anchoring Theory + Content, offered a brief comparison with IAT, which it builds upon, and used three examples to showcase IATC's capabilities. We showed that:

– IATC offers a more faithful representation of everyday mathematical practice than does, e.g., Lamport-style structured proof.

Fig. 16 Pie charts showing the relative proportion of IATC tags used to code MPM3, across five intuitive kinds of comments. E.g., Comment 1 has been categorised as a *Conjecture*. The IATC stanza `perf[assert](rel[stronger](rel[not](prove_rtf),  rel[not](random_test_false)))` associated with this comment (see Fig. 13) therefore adds these values to the usage counts within the *Conjecture* pie chart: '`Assert`' +1, '`stronger`' +1, and '`not`' +2. (Color figure online)
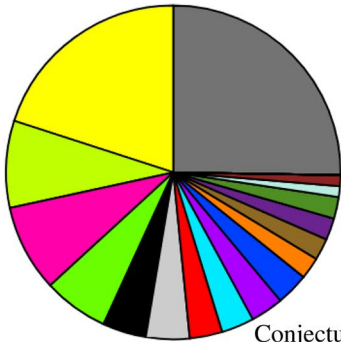
- IATC has the potential to support computational reasoning about mathematics by bringing structural relationships between pieces of mathematical content to the surface.
- IATC can recover salient elements of discourse within comments, as well as the way these contents connect across comments.

Some limitations to the approach should be considered when applying the framework. We emphasise that these are limitations and not necessarily flaws in the overall design. In general, the limitations could be addressed with extensions to the language.
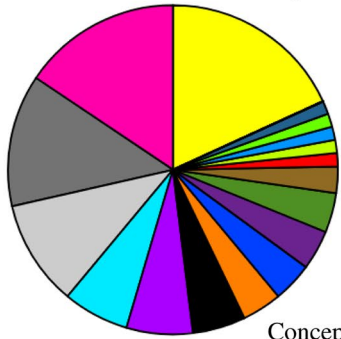
- IATC does not yet handle everything that is said in mathematical dialogues. We saw above that IATC nevertheless helps disambiguate the "other" category bracketed by Pease and Martin (2012).
- There are places where IATC representations remain bulky, pushing much of the actual reasoning into whatever representation system handles the content layer.
- One related limitation is that implications and assumptions that mathematicians consider "obvious" are typically elided from their discourse, often for valid expository reasons, and that, therefore, unpacking the contextual relationships between statements typically requires a mathematically trained annotator.
- We introduced a graphical way to segment dialogues, but IATC does not currently have the ability to express context shifts – although it can compare contexts with '`analogy`'.

Corneli et al. (2018) survey other relevant frameworks that might form extensions for a future version of IATC. More general-purpose formalisms like the W3C's "PROV" (Groth and Moreau 2013) would allow us to say something about the provenance and evolution of concepts, but would have nothing to say about the mathematics-specific features that interest us.
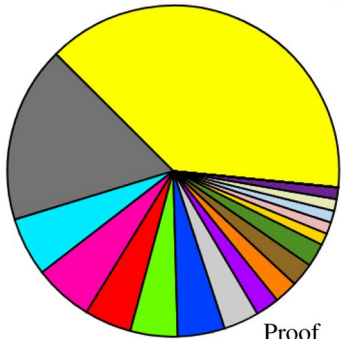
In Sect. 3, we mentioned that Discourse Representation Theory (DRT) has informed several earlier efforts to model mathematical discourse. We are aware of three PhD theses—by Clauss Zinn (2004), Mohan Ganesalingam (2013), and Marcos Cramer (2013)—which have made use of somewhat similar mathematics-specific interpretations of DRT. Zinn and Cramer focused on proof checking, while Ganesalingam looked at mathematical communication from a linguist's perspective. However, he opted to focus exclusively on mathematics in the "formal mode," leaving informal communication about matters such as "interestingness" to one side, because they bring with them a host of additional complications (Ganesalingam 2013, pp. 7–8). From a linguistic point of view, DRT is useful in a mathematical setting, in the first instance, because of its core ability to express "legitimate
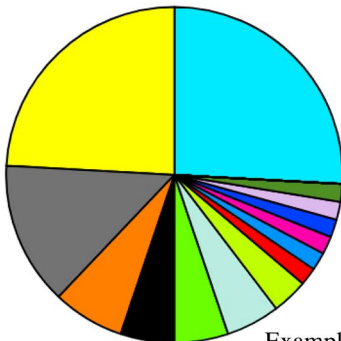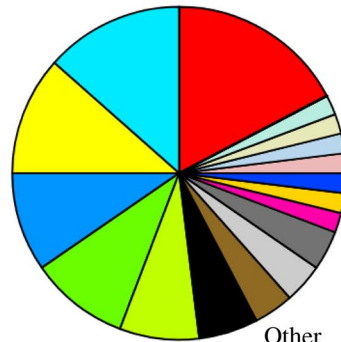
Agree
Assert
Challenge
Define
Judge
Query
Retract
Suggest
not
wlog
stronger
equivalent
has_property
case_split
instance_of
beautiful
easy
feasible
plausible
useful
auxiliary
goal
strategy
structural

Conjecture

Concept

Proof

Example

Other

antecedents for anaphor" (Ganesalingam 2013, p. 50). In Ganesalingam's work, this basic feature is extended to allow sidelong references to definite descriptions (such as 'the set of natural numbers') by "introducing generalised anaphors which can have presuppositional material attached to them" (Ganesalingam 2013, pp. 25, 237). Specifically, this allows one to infer from statements such as "x is prime" that x is in fact a member of the set of natural numbers (p. 25).

The associated requirement of combining semantics and pragmatics (van der Sandt 1992, p. 336) is reminiscent of our treatment of unspoken assertions and unstated features of content in our IATC-based analyses. To continue the comparison, Ganesalingam's adaptations of DRT overcame limitations, having to do with quantifier scoping, that constrained earlier type-theoretic analyses (Ganesalingam 2013, pp. 81–82). This is broadly similar to our use of nested structure in Sect. 4.1. Indeed, Sowa (2000) shows that several different approaches to nested structure (including DRT) are all mutually equivalent from a logical point of view. As indicated by van der Sandt (1992), pragmatics is relevant for DRT-based models because it can inform the context-specific resolution of Discourse Representation Schemes. This is related to the question we highlighted in Sect. 4.2: how to model the transitions between discourse moves in mathematics? IAT accounts for similar issues by making reference to dialogue norms, but we have seen that for mathematical dialogues, detailed content- and context-specific issues need to be taken into consideration at each stage. The models of content evolution used by Ganesalingam and Gowers (2017) to keep track of proof generation were structurally similar to the DRT-based models developed by Ganesalingam (2013): in this case, the evolution was governed by a limited set of reasoning tactics. Our work with IATC highlights features of mathematical reasoning, like `analogy`, that more general heuristics will need to account for.

There are other resources available which could further expand IATC's offerings in this regard. For example, a recent special issue of *Argument & Computation* (Harris and Marco 2017) includes papers detailing the usefulness of rhetorical structures for argument mining. Mitrović et al. (2017), in that volume, indicate the SALT Rhetorical Ontology (Groza 2012) as relevant prior work. SALT contains three categories—coherence relations, argument scheme relations, and rhetorical blocks—each of which unfolds with considerable further detail. These three categories can be seen as somewhat analogous to IATC's grammatical categories. Mitrović et al. (2017) and Lawrence et al. (2017) point to foundational work of Fahnestock (1999, 2004) on the argumentative function of rhetorical figures, particularly in science writing. IATC might be profitably connected to such analyses. Furthermore, the integration of rhetoric into argument mining highlights the relevance of structures that are rather different from the IAT-style transitions that have been used in work summarised by Budzynska et al. (2015). White's (1978, p. 6) pithy assertion that "logic itself is merely a formalization of tropical strategies" can serve as an additional provocation to develop structural analyses of this sort.

Nevertheless, whether mathematical content is modelled using ideas from logic, rhetoric, or other sources, considerable further work will be required to effectively describe the processes that are employed in forming and responding to mathematical arguments. A small case study included as an appendix to Pease et al. (2017)

(and, incidentally, based on MiniPolymath 3) illustrates the plausibility of Lakatos's model—however that model is clearly far from complete as a theory of mathematical production. Pease et al. were concerned with mathematical content only insofar as it fills slots for some 20 dialogue moves that are based on Lakatos's strategy for arguing about lemmas and counterexamples. For example, *MonsterBar*(m, c, r) gives a reason r, contradicting the justification m for the counter-conjecture not-c. At no point does this theory touch the supposed mathematical ground of axioms and rules of inference. That the reason r, for example, may have been formed inductively, or deductively, or in some other way, goes undiscussed. IATC would allow us to expand the structure that appears within statements like r. Whereas Pease et al.'s formalisation of Lakatosian reasoning as a dialogue game offers a computational model of certain dynamical patterns in mathematics, our current work has focused on kinematics. The efforts can be seen as complementary: Bundy (2013) has argued that the right representation can considerably simplify reasoning.

One promising approach to modelling process combines argumentation and multi-agent systems (Modgil and McGinnis 2007; Maghraby et al. 2012; Robertson 2012). However, most approaches to modelling specifically *mathematical* agents have had significant limitations. Thus, for example, Fiedler and Horacek (2007) have described the difficulty of squaring argumentation-theoretic work with the methods of formal proof. Ganesalingam and Gowers's (2017) project aimed at simulating a solitary individual rather than a population. However, Furse (1990) had already called into question the robustness of approaches to modelling mathematical creativity that only model a solitary creative individual. Pease et al. (2009) describe an implementation effort that made use of a multi-agent approach, drawing on argumentation theory concepts and a Lakatosian model of dialogue. However, the mathematical applications of that system were limited to straightforward computational aspects of number theory and group theory, which suggests a "knowledge bottleneck" (Saint-Dizier 2016; Moens 2018).

As indicated in a report of the National Research Council (2014, p. 90), "knowledge extraction and structuring in the context of mathematics" is in demand on an increasingly industrial scale. IATC allows methods of argumentation to interface with those of knowledge representation; both aspects are relevant to knowledge extraction. Formalisation of IATC would assist in its applicability: "IKL Conceptual Graphs" defined by Sowa (2008) would provide a natural foundation. IKL, the *IKRIS Knowledge Language* (Hayes 2006; Sowa 2008), deals elegantly with context and has been used as a representational formalism in a project with aims comparable to our own: the Slate project (Bringsjord et al. 2008), which centred on an argumentation tool that could support a mixture of deductive and informal reasoning.[4] Previous work on mathematical usage can also inform future efforts in knowledge modelling with IATC (Trzeciak 2012; Wells 2003; Wolska 2015; Ginev 2011).

Mathematical Knowledge Management, particularly in the "flexiformal" understanding developed by Kohlhase (2012) and Kohlhase et al. (2017), presents another paradigm that could eventually be integrated with IATC.

---

[4] http://www.jfsowa.com/ikl/IKLslate.pdf.

Flexiformality combines strict formalisations of those parts of mathematics for which that makes sense with opaque representations of constants, objects, and informal theories. Iancu (2017) built on Kohlhase's work, and focused on "co-representing both the narration and content aspects of mathematical knowledge in a structure preserving way" (pp. 3–4). However, modelling narrative in Iancu's sense is more relevant to the "frontstage" presentation of mathematics in a single authorial voice than to the "backstage" production of mathematics (cf. Hersh 1991)). Section 4.2 illustrated one such example from backstage: mathematicians need to be able to choose between different mathematical problems.

IATC offers a step forward for research into both the communication and production of mathematics, and can play a role in future work on knowledge extraction and simulation. Potential applications include, among others, the development of a new generation of mathematics tutoring software and digital assistants that engage their users in thought-provoking dialogues.

Appendix: Reference Coding Samples

This "Appendix" collects sample texts and IATC codings to supplement Tables 1 and 2 in Sect. 3, which introduced the available codes. Texts are sourced from the examples discussed in Sect. 4. In general, one utterance may expand to multiple statements in IATC; accordingly, texts may appear here multiple times. Bold face is used to illustrate the portion of the text, at right, that justifies the tag that appears, at left. Numbering refers to the tree-ordering of MiniPolymath 3 comments, unless another source is indicated.

Performatives	
perf[**assert**](rel[implies](rel[not](prove_rtf), rel[not](random_test_false)))	1. Could you start off with a random point in the plane and prove it doesn't work, **if you can't prove that then the opposite holds**
perf[**agree**](cycle_partition)	2.2.1.1. **I believe this is true.** It proves that it's enough to find a cycle that visits each vertex at least once

Performatives	
perf[**challenge**](problem, equi_tri_stuck)	3.1.1. 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
perf[**retract**](perf[challenge](problem, equi_tri_stuck))	3.1.1.2.1. **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**
perf[**define**](pivot_seq, ps_def)	2. 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**
perf[**suggest**](meta[goal](cycle_spans_S))	2.1. Nice. We need only to consider the times when two points are connected—this gives us a path, and after some time this path will come back to some already visited point. So there is a cycle. **If only we could find a cycle which spans all the points, the question is solved**. That may be some useful simplification
perf[**judge**](value[useful](pivot_seq))	2.1. Nice. We need only to consider the times when two points are connected—this gives us a path, and after some time this path will come back to some already visited point. So there is a cycle. If only we could find a cycle which spans all the points, the question is solved. **That may be some useful simplification**
perf[**query**](random_test_false)	1. **Could you start off with a random point in the plane and prove it doesn't work**, if you can't prove that then the opposite holds
perf[**queryE**](additional_condition_on cycles(X))	2.1.1.1. For example, the restriction on how the next pivot is chosen (geometrically: comment 9). **Are there any other restrictions?** Can we start with a complete graph and all cycles on that graph and just discard the ones that don't follow the restrictions to converge on the ones that do?

Inferential structure	
perf[assert](rel[**implies**](rel[not](prove_rtf), rel[not](random_test_false)))	1. Could you start off with a random point in the plane and prove it doesn't work, **if you can't prove that then the opposite holds**
perf[assert](rel[**equivalent**](problem, forall_exists_problem), cycle_partition)	2.2.1.1. I believe this is true. It proves that **it's enough to find a cycle that visits each vertex at least once**. There are no "rho" processes with an initial segment that doesn't repeat

Inferential structure	
perf[assert](rel[implies](rel[not](prove_rtf, rel[**not**](random_test_false)))	1. Could you start off with a random point in the plane and **prove it doesn't work**, if you can't prove that then the opposite holds
perf[question](rel[implies](rel[**conjunction**] (G_infinite_group, H_subgroup_of_G, H_ finite_index_in_G), G_not_equal_to_union_of_ cosets))	(Sect. 4.2, Question) I have seen this problem, that if G **is a finite group and H is a proper subgroup of G with finite index** then $G \neq \cup\_g \in GgHg^{-1}$. Does this remain true for the infinite case also?
perf[assert](rel[**has_property**](pivot_seq, has_cycle))	2.1. Nice. We need only to consider the times when two points are connected—**this gives us a path, and after some time this path will come back to some already visited point. So there is a cycle.** If only we could find a cycle which spans all the points, the question is solved. That may be some useful simplification
rel[**instance_of**](S, convex_plus_point)	3.1. Yes. Can we **do it if there is a single point not on the convex hull of the points?**
perf[assert](rel[**indep_of**](disj_path, M))	(MPM1, 31.) Quick thought following on David Speyer's first comment: The problem asks us to prove that no set of size $(n - 1)$ can disconnect two diagonally opposing vertices in the n-cube. By Menger's theorem, this is equivalent to proving that there are n internally vertex-disjoint paths between these two vertices. So, now we are faced with **a constructive problem, independent of the set M: Construct n vertex-disjoint paths from 0^n to 1^n in the n-cube**
perf[assert](rel[**case_split**](IS, IS_A, IS_B))	2.1.1.1.1. The line must sweep out a full rotation (and only one full rotation) of 2π during the traversal of S. I feel like this is intimately related to proving that there is a starting angle for any point P in S such that all of S is then traversed. I'm trying to show this by induction. Base case ($|S| = 2$) is obvious. **Let $|S| = n$, take $S' = S \cup \{Q\}$, and start with some windmill traversal of S. Case A: Q is unreachable. Therefore we just traverse S, taking 2π to do so by induction. Case B: Q is reachable at some angle. [...]**
perf[assert](rel[**wlog**](problem, zero_angle), one_turn)	11.2.3.1. Only the starting point matters. By the problem statement, it appears that the initial angle is irrelevant to the existence of a pivot point P^* from which all of S is traversed. Every point in S is a pivot point, but only with a specific range of starting angle (e.g. those consistent with the cycle generating S). The union of these intervals must necessarily be $[0, 2\pi)$, and thus **we can assume WLOG that the starting angle is 0** (and thus we single out a specific point—or points in the case of $|S| = 2$)

Heuristic value judgments	
perf[judge](value[**easy**](S_is_conv))	3. **If the points form a convex polygon, it is easy**
perf[judge](rel[not](value[**plausible**](equi_tri_stuck)))	3.1.1. 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**
perf[judge](value[**beautiful**](proof_sugg))	14.2.1. **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?
perf[judge](value[**useful**](pivot_seq))	2.1. Nice. We need only to consider the times when two points are connected—this gives us a path, and after some time this path will come back to some already visited point. So there is a cycle. **If only we could find a cycle which spans all the points, the question is solved. That may be some useful simplification**

Reasoning tactics	
perf[suggest](meta[**goal**](cycle_spans_S))	2.1. Nice. We need only to consider the times when two points are connected—this gives us a path, and after some time this path will come back to some already visited point. So there is a cycle. **If only we could find a cycle which spans all the points, the question is solved**. That may be some useful simplification
perf[suggest](meta[**strategy**](cycle_spans_S, process_of_elim))	2.1.1.1. For example, the restriction on how the next pivot is chosen (geometrically: comment 9). Are there any other restrictions? **Can we start with a complete graph and all cycles on that graph and just discard the ones that don't follow the restrictions to converge on the ones that do?**
perf[suggest](meta[**auxiliary**](problem, forall_split))	Ok. I think the solution might involve this observation, **with the observation that every point participates in a "splitting" line (one with $n/2$ points on one side)**
perf[assert](meta[**analogy**](compute 500th digit of (sqrt(2)+sqrt(3))^2012, compute 500th digit of (x+y)^2012))	(Sect. 4.1) **Can we do this for $x + y$? For e? Rationals with small denominator?**
perf[assert](rel[**implements**](#SUBGRAPH, the trick might be))	(Sect. 4.1) $(\sqrt{2} + \sqrt{3})^{2012} + (\sqrt{3} - \sqrt{2})^{2012}$ **is an integer! And** $(\sqrt{3} - \sqrt{2})^{2012}$ **is a very small number.** Maybe the final answer is "9"?
rel[**generalise**](binomial theorem, eliminate cross terms)	(Sect. 4.1) $(\sqrt{2} + \sqrt{3})^{2012} + (\sqrt{2} - \sqrt{3})^{2012}$ **is an integer!**

Content-focused structural relations	
rel[**used_in**](pivot_seq, pivot)	2. 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
rel[**reform**](H_finite_index_in_G , H_infinite_index_in_G)	(Sect. 4.2, Second comment on question) Yes, the statement is out of focus: gHg^{-1} is intended (**and "infinite index case"**). The natural starting point is to ask whether the proof for finite index breaks down
rel[**instantiates**]((sqrt(2) + sqrt(3))^2,general form of the problem)	(Sect. 4.1) $(\sqrt{2} + \sqrt{3})^2$
perf[assert](rel[**expands**](2 + 2sqrt2sqrt3 + 3, (sqrt(2)+sqrt(3))^2))	(Sect. 4.1) $2 + 2\sqrt{2}\sqrt{3} + 3$
perf[assert](rel[**sums**](2 + 2sqrt(2)sqrt(3) + 2 + 2 - 2sqrt(2)sqrt(3)+3,10))	(Sect. 4.1) $(\sqrt{2} + \sqrt{3})^2 + (\sqrt{2} - \sqrt{3})^2 = 10$
perf[assert](rel[**cont_summand**]((sqrt(2) + sqrt(3))^2012 + (sqrt(2)-sqrt(3))^2012,(sqrt(3)-sqrt(2))^2012)	(Sect. 4.1) And $(\sqrt{3} - \sqrt{2})^{2012}$ is a very small number. Maybe the final answer is "9"?

References

Aberdein, A., and I.J. Dove. 2013. Introduction. In *The argument of mathematics*, ed. A. Aberdein and I.F. Dove, 1–8. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-6534-4_1.

Aliseda, A. 2003. Mathematical reasoning versus abductive reasoning: a structural approach. *Synthese* 134(1–2): 25–44.

Barany, M. 2010. '[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: Ayers P, Ortega F (eds) Proceedings of the 6th international symposium on Wikis and open collaboration, ACM.

Botting, D. 2015. Inferences and illocutions. *Argument & Computation* 6(3): 246–264.

Bringsjord, S., J. Taylor, A. Shilliday, M. Clark, K. Arkoudas, M. Schoelles, M. Destefano, and J. Wodicka. 2008. Slate: an argument-centered intelligent assistant to human reasoners. In Grasso F, Green N, Kibble R, and Reed C (eds) Proceedings of the 8th international workshop on computational models of natural argument (CMNA 2008), pp 1–10.

Budzynska, K. 2013. Circularity in ethotic structures. *Synthese* 190(15): 3185–3207.

Budzynska, K., and C. Reed. 2011. Speech acts of argumentation: inference anchors and peripheral cues in dialogue. In Grasso F, Green N, and Reed C (eds) Computational models of natural argument: papers from the 2011 AAAI workshop. http://www.aaai.org/ocs/index.php/WS/AAAIW11/paper/view/3940.

Budzynska, K., M. Janier, C. Reed, and P. Saint-Dizier. 2013. Towards extraction of dialogical arguments. In Grasso F, Green N, and Reed C (eds) Proceedings of 13th international conference on computational models of natural argument (CMNA 2013).

Budzynska, K., M. Janier, J. Kang, C. Reed, P. Saint-Dizier, M. Stede, and O. Yaskorska. 2014a. Towards argument mining from dialogue. In Parsons S, Oren N, Reed C, and Cerutti F (eds) Computational models of argument: proceedings of COMMA 2014, IOS Press, Frontiers in artificial intelligence and applications, vol 266, pp 185–196. http://comma2014.arg.dundee.ac.uk/res/pdfs/19-budzynska.pdf.

Budzynska, K., M. Janier, C. Reed, P. Saint-Dizier, M. Stede, and O. Yaskorska. 2014b. A model for processing illocutionary structures and argumentation in debates. In Calzolari N, Choukri K, Declerck T, Loftsson H, Maegaard B, Mariani J, Moreno A, Odijk J, and Piperidis S (eds) Proceedings of the

ninth international conference on language resources and evaluation (LREC-2014), Reykjavik, Iceland, May 26–31, 2014, pp 917–924. http://www.lrec-conf.org/proceedings/lrec2014/summaries/77.html.

Budzynska, K., M. Janier, J. Kang, B. Konat, C. Reed, P. Saint-Dizier, M. Stede, and O. Yaskorska. 2015. Automatically identifying transitions between locutions in dialogue. In Mohammed D and Lewinski M (eds) Argumentation and reasoned action: proceedings of the 1st European conference on argumentation, Lisbon 2015, Volume II, College Publications, no. 63 in Studies in Logic and Argumentation, pp 311–328.

Budzynska, K., M. Janier, C. Reed, and P. Saint-Dizier. 2016. Theoretical foundations for illocutionary structure parsing. *Argument & Computation* 7(1): 91–108.

Bundy, A. 1988. The use of explicit plans to guide inductive proofs. In Lusk E, Overbeek R (eds) 9th International conference on automated deduction, Argonne, Illinois, USA, May 23–26, 1988. Proceedings, Springer, pp 111–120.

Bundy, A. 2013. The interaction of representation and reasoning. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 469(2157), http://rspa.royalsocietypublishing.org/content/469/2157/20130194.

Carrascal, B. 2015. Proofs, mathematical practice and argumentation. *Argumentation* 29(3): 305–324. https://doi.org/10.1007/s10503-014-9344-0.

Chandrasekhar, et al. 2010. Group cannot be the union of conjugates. http://mathoverflow.net/q/34044.

Corneli, J., U. Martin, D. Murray-Rust, and A. Pease. 2017a. Towards mathematical AI via a model of the content and process of mathematical question and answer dialogues. In Geuvers H, England M, Hasan O, Rabe F, and Teschke O (eds) Intelligent computer mathematics 10th international conference, CICM 2017, Edinburgh, UK, 2017, Proceedings.

Corneli, J., U. Martin, D. Murray-Rust, A. Pease, R. Puzio, and G. Rino Nesin. 2017b. Modelling the way mathematics is actually done. In Sperber M, Bresson J, Santolucito M, and McLean A (eds) 2017 International Workshop on functional art, music, modelling and design (FARM 2017), ACM.

Corneli, J., A. Pease, and D. Stefanou. 2018. Social aspects of concept invention. In *Concept invention: foundations, implementation, social aspects and applications*, ed. R. Confalonieri, A. Pease, M. Schorlemmer, T. Besold, O. Kutz, E. Maclean, and M. Kaliakatsos-Papakostas. Computational Synthesis and Creative Systems: Springer.

Cramer, M. 2013. Proof-checking mathematical texts in controlled natural language. Ph.D. thesis, Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn.

Dauphin, J., and M. Cramer. 2018. ASPIC-END: structured argumentation with explanations and natural deduction. In Black E, Modgil S, and Oren N (eds) Theory and applications of formal argumentation 4th international workshop, TAFA 2017, Melbourne, VIC, Australia, August 19–20, 2017, Revised Selected Papers, Springer, no. 10757 in LNAI, pp 51–66.

Dove, I.J. 2009. Towards a theory of mathematical argument. *Foundations of Science* 14(1–2): 137–152.

Dufour, M. 2013. Arguing around mathematical proofs. In *The argument of mathematics*, ed. A. Aberdein and I.J. Dove, 61–76. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-6534-4_5.

Dutilh Novaes, C. 2016. Reductio ad absurdum from a dialogical perspective. *Philosophical Studies* 173(10): 2605–2628.

Epstein, S.L. 2015. Wanted: collaborative intelligence. *Artificial Intelligence* 221: 36–45.

Fahnestock, J. 1999. *Rhetorical figures in science*. Oxford: Oxford University Press.

Fahnestock, J. 2004. Figures of argument (OSSA 2005 Keynote Address). Informal Logic 24(2).

Fiedler, A., and H. Horacek. 2007. Argumentation within deductive reasoning. *International Journal of Intelligent Systems* 22(1): 49–70.

Furse, E. 1990. Why did AM run out of steam? Tech. rep., CS-90-4, Department of Computer Studies, University of Glamorgan.

Ganesalingam, M. 2013. *The language of mathematics: a linguistic and philosophical investigation*, LNCS, vol 7805. Springer.

Ganesalingam, M., and W. Gowers. 2017. A fully automatic theorem prover with human-style output. *Journal of Automated Reasoning* 58: 253–291.

Gasteren, A.J. 1990. *On the shape of mathematical arguments*, vol. 445., Lecture Notes in Computer Science Berlin: Springer.

Ginev, D. 2011. The structure of mathematical expressions. Master's thesis. Bremen, Germany: Jacobs University.

Gowers, W. 2017. How do human mathematicians avoid big searches? https://www.newton.ac.uk/seminar/20170728133014301, Lecture, Fri, July 28, 13:30–14:30, Isaac Newton Institute, University of Cambridge.

Gowers, W., and M. Ganesalingam. 2012. Modelling the mathematical discovery process. Maxwell Institute Lecture, Fri, November 2, 4pm – 5pm, James Clerk Maxwell Building, University of Edinburgh.

Groth, P., and L. Moreau. 2013. PROV-overview. An overview of the PROV family of documents. In World Wide Web consortium. https://www.w3.org/TR/prov-overview/.

Groza, T. 2012. *Advances in semantic authoring and publishing, Studies on the Semantic Web*, vol. 13. Heidelberg: IOS Press.

Harris, R., and C. D. Marco. 2017. Argument & Computation 8(3) [Special Issue on Rhetorical figures, arguments, computation.]

Hayes, P. 2006. IKL guide. http://www.ihmc.us/users/phayes/IKL/GUIDE/GUIDE.html.

Hendrix, GG. 1975. Partitioned networks for the mathematical modeling of natural language semantics. Ph.D. thesis, The University of Texas at Austin, also archived as Tech. Report NL-28, Department of Computer ScienceD.

Hendrix, G. G. 1979. Encoding knowledge in partitioned networks. In *Associative networks: representation and use of knowledge by computers*, ed. N. V. Findler, 51–92. New York: Academic Press.

Hersh, R. 1991. Mathematics has a front and a back. *Synthese* 88(2): 127–133.

Iancu, M. 2017. Towards flexiformal mathematics. Ph.D. thesis, Jacobs University, Bremen, DE.

Inglis, M., J.P. Mejia-Ramos, and A. Simpson. 2007. Modelling mathematical argumentation: the importance of qualification. *Educational Studies in Mathematics* 66(1): 3–21.

Janier, M., and C. Reed. 2017. Towards a theory of close analysis for dispute mediation discourse. *Argumentation* 31(1): 45–82.

Kamp, H., and U. Reyle. 1993. *From discourse to logic introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory, studies in linguistics and philosophy*, vol 42. Springer.

Karttunen, L. 1976. Discourse referents. In *Syntax and semantics*, vol 7, ed. J.D. McCawley. Notes from the Linguistic Underground, Academic Press, pp 363–386D.

Klaus, K. 2004. *Content analysis: an introduction to its methodology*, 2nd ed. Thousand Oaks: Sage Publications.

Kohlhase, M. 2012. The flexiformalist manifesto. In Voronkov A, Negru V, Ida T, Jebelean T, Petcu D, Watt S, and Zaharie D (eds) Symbolic and numeric algorithms for scientific computing (SYNASC 2012), 14th international symposium on, IEEE, pp 30–35.

Kohlhase, M., T. Koprucki, D. Müller, and K. Tabelow. 2017. Mathematical models as research data via flexiformal theory graphs. In Geuvers H, England M, Hasan O, Rabe F, and Teschke O (eds) Intelligent computer mathematics 10th international conference, CICM 2017, Edinburgh, UK, July 17–21, 2017, Proceedings, Springer, LNAI, vol 10383, pp 224–238.

Lakatos, I. 1976. *Proofs and refutations: the logic of mathematical discovery*. Cambridge: Cambridge University Press.

Lamport, L. 1995. How to write a proof. *The American Mathematical Monthly* 102(7): 600–608.

Lamport, L. 1999. Specifying concurrent systems with TLA$^+$. *NATO Science Series, III: Computer and Systems Sciences* 173(173): 183–247.

Lamport, L. 2012. How to write a 21st century proof. *Journal of Fixed Point Theory and Applications* 11(1): 43–63.

Lamport, L. 2014. TLA^{+2}: *a preliminary guide*. http://lamport.azurewebsites.net/tla/tla2-guide.pdf.

Lamport, L. 2015. *The* TLA$^+$*hyperbook*. http://lamport.azurewebsites.net/tla/hyperbook.html.

Larvor, B. 2012. How to think about informal proofs. *Synthese* 187(2): 715–730.

Lawrence, J., F. Bex, C. Reed, and M. Snaith. 2012. AIFdb: infrastructure for the argument web. In Verheij B, Szeider S, and Woltran S (eds) Computational models of argument: proceedings of COMMA 2012, IOS Press, Frontiers in artificial intelligence and applications, vol 245, pp 515–516.

Lawrence, J., J. Visser, and C. Reed. 2017. Harnessing rhetorical figures for argument mining. *Argument & Computation* 8(3): 289–310.

Maghraby, A., D. Robertson, A. Grando, and M. Rovatsos. 2012. Automated deployment of argumentation protocols. In Verheij B, Szeider S, and Woltran S (eds) Computational models of argument: proceedings of COMMA 2012, IOS Press, Frontiers in artificial intelligence and applications, vol 245, pp 197–204.

Martin, U. 2015. Stumbling around in the dark: lessons from everyday mathematics. In *Automated deduction—CADE-25*, ed. A. P. Felty and A. Middeldorp, 29–51. Cham: Springer. https://doi.org/10.1007/978-3-319-21401-6_2.

Martin, U., and A. Pease. 2013. What does mathoverflow tell us about the production of mathematics? In Novak J and Jaimes A (eds) SOHUMAN, 2nd international workshop on social media for crowdsourcing and human computation, at ACM Web Science 2013, May 1, 2013, Paris. https://arxiv.org/abs/1305.0904.

Mercier, H., and D. Sperber. 2011. Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences* 34(2): 57–74.

Mitrović, J., C. OReilly, M. Mladenović, and S. Handschuh. 2017. Ontological representations of rhetorical figures for argument mining. *Argument & Computation* 8(3): 267–287.

Modgil, S., and J. McGinnis. 2007. Towards characterising argumentation based dialogue in the argument interchange format. In Argumentation in multi-agent systems: 4th international workshop, ArgMAS 2007, Honolulu, HI, USA, May 15, 2007, ed. I. Rahwan, S. Parsons, and C. Reed, 80–93. Revised Selected and Invited Papers: Springer.

Moens, M. F. 2018. Argumentation mining: How can a machine acquire common sense and world knowledge? Argument & Computation 9(1).

National Research Council. 2014. *Developing a 21st century global library for mathematics research*. National Academies Press.

Nielsen, M. et al. 2009–2018. Polymath wiki. http://michaelnielsen.org/polymath1/index.php?title=Main_Page.

Pease, A., and U. Martin. 2012. Seventy four minutes of mathematics: an analysis of the third Mini-Polymath project. In Larvor B and Pease A (eds) Proceedings of AISB/IACAP 2012, symposium on mathematical practice and cognition II. http://homepages.inf.ed.ac.uk/apease/papers/seventy-four.pdf.

Pease, A., A. Smaill, S. Colton, and J. Lee. 2009. Bridging the gap between argumentation theory and the philosophy of mathematics. *Foundations of Science* 14(1): 111–135.

Pease, A., J. Lawrence, K. Budzynska, J. Corneli, and C. Reed. 2017. Lakatos-style collaborative mathematics through dialectical, structured and abstract argumentation. Artificial Intelligence 246:181–219. http://www.sciencedirect.com/science/article/pii/S0004370217300267.

Pedemonte, B. 2007. How can the relationship between argumentation and proof be analysed? *Educational Studies in Mathematics* 66 (1): 23–41. https://doi.org/10.1007/s10649-006-9057-x.

Reed, C., and K. Budzynska. 2010. How dialogues create arguments. In van Eemeren F, Garssen B, Godden D and Mitchell G (eds) Proceedings of the 7th conference of the international society for the study of argumentation (ISSA 2010).

Reed, C., K. Budzynska, R. Duthie, M. Janier, B. Konat, J. Lawrence, A. Pease, and M. Snaith. 2017. The argument web: an online ecosystem of tools, systems and services for argumentation. *Philosophy & Technology* 30(2): 137–160. https://doi.org/10.1007/s13347-017-0260-8.

Robertson, D. 2012. Lightweight coordination calculus for agent systems: retrospective and prospective. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 84–89. https://doi.org/10.1007/978-3-642-29113-5_7.

Robinson, J.A. 1965. A machine-oriented logic based on the resolution principle. *Journal of the ACM (JACM)* 12(1): 23–41.

Saint-Dizier, P. 2016. Argument mining: the bottleneck of knowledge and language resources. In Calzolari N, Choukri K, Declerck T, Goggi S, Grobelnik M, Maegaard B, Mariani J, Mazo H, Moreno A, Odijk J, and Piperidis S (eds) 10th International conference on language resources and evaluation (LREC 2016), pp 983–990.

Snaith, M., and C. Reed. 2012. TOAST: online ASPIC+ implementation. In Verheij B, Szeider S, and Woltran S (eds) Computational models of argument: proceedings of COMMA 2012, IOS Press, Frontiers in artificial intelligence and applications, vol 245, pp 509–510.

Snaith, M., and C. Reed. 2016. Dialogue grammar induction. In Mohammed D and Lewinski M (eds) Argumentation and reasoned action, volume I: proceedings of the 1st European conference on argumentation, Lisbon 2015, Studies in Logic and Argumentation, vol 62, College Publications.

Snaith, M., J. Devereux, J. Lawrence, and C. Reed. 2010. Pipelining argumentation technologies. In Baroni P, Cerutti F, Giacomin M, and R S. G. (eds) Computational models of argument: proceedings of COMMA 2010, IOS Press, Frontiers in artificial intelligence and applications, vol 216, pp 447–454.

Sowa, J. F. 2000. *Knowledge representation: logical, philosophical, and computational foundations*. MIT Press

Sowa, J. F. 2008. Conceptual graphs. In *Handbook of knowledge representation*, ed. van Harmelen F, Lifschitz V, Porter B, chap 5, pp 213–237. Elsevier.

Sowa, J. F., and A. K. Majumdar. 2003. Analogical reasoning. In Aldo A, Lex W, and Ganter B (eds) Conceptual structures for knowledge creation and communication: 11th international conference on conceptual structures, ICCS 2003, Dresden, Germany, July 21–25, 2003, Proceedings, Springer, no. 2746 in LNAI, pp 16–36.

Tanswell, F. 2015. A problem with the dependence of informal proofs on formal proofs. *Philosophia Mathematica* 23 (3): 295. https://doi.org/10.1093/philmat/nkv008.

Tao, T., et al. 2009. IMO 2009 Q6 as a mini-polymath project. https://wp.me/p3qzP-Ef.

Tao, T., et al. 2011. Minipolymath3 project: 2011 IMO. https://wp.me/pAG2F-41.

Trzeciak, J. 2012. *Mathematical English usage. A dictionary*. http://www.emis.de/monographs/Trzeciak/biglist.html.

van der Sandt, R.A. 1992. Presupposition projection as Anaphora resolution. *Journal of Semantics* 9(4): 333–377. https://doi.org/10.1093/jos/9.4.333.

van Oers, B. 2002. Fruits of polyphony: a commentary on a multiperspective analysis of mathematical discourse. *Journal of the Learning Sciences* 11(2–3): 359–363. https://doi.org/10.1080/10508406.2002.9672143.

Visser, J., F. Bex, C. Reed, and B. Garssen. 2011. Correspondence between the pragma-dialectical discussion model and the argument interchange format. *Studies in Logic, Grammar and Rhetoric* 23(36): 189–224.

Walton, D., C. Reed, and F. Macagno. 2008. *Argumentation schemes*. New York: Cambridge University Press.

Wells, C. 2003. *A handbook of mathematical discourse*. West Conshohocken: Infinity Publishing.

White, H.V. 1978. *Tropics of discourse: essays in cultural criticism*. Baltimore: Johns Hopkins University Press.

Wolska, M. A. 2015. Students' language in computer-assisted tutoring of mathematical proofs. No. 40 in Saarbrücken Dissertations in Language Science and Technology, Universaar.

Zack, V., and B. Graves. 2001. Making mathematical meaning through dialogue: "once you think of it, the Z minus three seems pretty weird". *Educational Studies in Mathematics* 46(1): 229. https://doi.org/10.1023/A:1014045408753.

Zinn, C. 2004. Understanding informal mathematical discourse. Ph.D. thesis, Institut fur Informatik, Universitat Erlangen-Nurnberg.