



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Using crowdsourced mathematics to understand mathematical practice

Citation for published version:

Pease, A, Martin, U, Tanswell, FS & Aberdein, A 2020, 'Using crowdsourced mathematics to understand mathematical practice', *ZDM*, vol. 52, no. 6, pp. 1087-1098. <https://doi.org/10.1007/s11858-020-01181-7>

Digital Object Identifier (DOI):

[10.1007/s11858-020-01181-7](https://doi.org/10.1007/s11858-020-01181-7)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

ZDM

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





Using crowdsourced mathematics to understand mathematical practice

Alison Pease¹ · Ursula Martin² · Fenner Stanley Tanswell³ · Andrew Aberdein⁴

Accepted: 5 July 2020 / Published online: 18 July 2020
© The Author(s) 2020

Abstract

Records of online collaborative mathematical activity provide us with a novel, rich, searchable, accessible and sizeable source of data for empirical investigations into mathematical practice. In this paper we discuss how the resources of crowdsourced mathematics can be used to help formulate and answer questions about mathematical practice, and what their limitations might be. We describe quantitative approaches to studying crowdsourced mathematics, reviewing work from cognitive history (comparing individual and collaborative proofs); social psychology (on the prospects for a measure of collective intelligence); human–computer interaction (on the factors that led to the success of one such project); network analysis (on the differences between collaborations on open research problems and known-but-hard problems); and argumentation theory (on modelling the argument structures of online collaborations). We also give an overview of qualitative approaches, reviewing work from empirical philosophy (on explanation in crowdsourced mathematics); sociology of scientific knowledge (on conventions and conversations in online mathematics); and ethnography (on contrasting conceptions of collaboration). We suggest how these diverse methods can be applied to crowdsourced mathematics and when each might be appropriate.

1 Introduction

While mathematicians have collaborated since antiquity, online collaborations among large numbers of mathematicians are a novelty in terms of scale, speed, anonymity, and transparency. They provide new opportunities for the practice of mathematics, and thereby for scholars of that practice in general, and mathematics education researchers in particular. Many mathematics educators agree that students should be exposed to the practices of working mathematicians, even if they do not always agree what those practices are (Stillman et al. 2020). Records of online collaborative activity form temptingly accessible, novel, rich, searchable and sizeable sources of data. Unsurprisingly, scholars of mathematical practice have started to look to such sources for insights we should expect to impact mathematics education. Specifically, much mathematical practice research

either lacks access to the knowledge generation process, as with historical studies, or is at some level artificial, as with laboratory studies of mathematicians. Crowdsourcing addresses both limitations.¹

In this paper we ask three questions: *What can we learn from these new socio-technical projects? How does crowdsourced mathematics differ from traditional mathematics? How can studies of crowdsourcing be used to better understand mathematical practice in general?*

Here are three preliminary sets of answers. The strongest is that there is no difference *in kind* between crowdsourced mathematics and traditional mathematics, so one can inform us about the other. Some mathematicians' reflections on crowdsourcing support this: for instance, in highlighting speed, Terry Tao draws a difference of degree rather than kind.² We may also compare crowdsourced proofs to proofs constructed traditionally, whether by comparing reviewers'

✉ Andrew Aberdein
aberdein@fit.edu

¹ University of Dundee, Dundee, UK

² University of Edinburgh, Edinburgh, UK

³ Loughborough University, Loughborough, UK

⁴ Florida Institute of Technology, Melbourne, USA

¹ In the study of mathematical practices, there is a substantial difficulty of access: many salient features of those practices are not readily accessible to be studied. As such, even for those not interested in crowdsourcing per se, the studies described in this paper are still important in allowing a different means of accessing and researching mathematical practices.

² “If mathematicians had been attacking the problem in the standard way, with what [Tao] describes as a ‘flood of mini-papers’, it might have taken years to get the bound down that far” (Ball 2014, p. 422).

reports, constructing and comparing argument maps, analysis of ‘fruitfulness’—the number of new lemmas, concepts and examples involved—or otherwise.

A more speculative answer is that although crowdsourced mathematics is a new practice, it is rapidly becoming mainstream. The internet is now part of traditional mathematical practice: essential to core knowledge production, not just a fringe novelty or mere facilitator of administrative tasks.

A third, more moderate answer is that although crowdsourced mathematics differs from traditional mathematics, it is an interesting mathematical practice in its own right. Indeed, we might query the existence of “traditional mathematics”, arguing that any claim about or investigation into mathematical practice must reflect a specific mathematical culture and socio-historical context. Mathematical practice has changed since ancient times: technological inventions, distribution mechanisms, changing applications, geographical and linguistic variations, social and institutional differences, historical understanding of the world and our place in it, developments in representation and notation and mathematical inventions themselves, all change the landscape of mathematical research in ways which are rarely tested. The job of a mathematician has changed significantly within living memory, so some claims about mathematical practice made within living memory no longer hold. If “traditional mathematics” does not exist, then researchers into crowdsourced mathematics cannot say how their work relates to it, but can still rightly claim to study an important mathematical practice (and how it relates to other such practices).

Rather than argue prematurely for any of these three approaches, we survey methods, disciplines, and research questions drawn from each of them. Hence we will consider both studies that explore how crowdsourced mathematics differs from traditional mathematics and studies that treat it as a representative sample of wider practices. We then make recommendations for future scholars of crowdsourced mathematics and conclude by considering some limitations of such studies.

2 Crowdsourcing

Before we can learn about mathematical practices from crowdsourced mathematics, we should understand the term ‘crowdsourcing’. Crowdsourcing presumes that a crowd can solve problems or complete tasks better than an individual or group of individuals can, for various meanings of ‘better’ such as faster, more efficiently, more creatively, or just

being able to solve the problem at all. The idea is popularly known as the “wisdom of crowds”, after Galton’s (1907) observation that the mean of the guesses of the weight of an ox at a fair was within one pound of the actual weight (for more detail, see Surowiecki 2004). In more general terms, the wisdom of crowds effect extends far beyond averaging or voting: for example, *Galaxy Zoo* draws on the crowd to classify large numbers of telescope images of distant galaxies, and to collectively identify particularly interesting ones (Raddick et al. 2010).

Nonetheless, it is easy to find counterexamples, cases where crowds are worse in some way than individuals would be, such as mob justice, panic buying, or online anti-science groups. This can be for a range of reasons, such as groups compounding their prejudices, echo chamber effects, feedback loops, game-theoretic strategies emerging, social norms, or intentional sabotage. Therefore, crowdsourcing projects must consider what features of a crowd are salient in securing crowdsourcing’s benefits and avoiding potential drawbacks.

In their systematic review of definitions of crowdsourcing, Estellés-Arolas and González-Ladrón-de-Guevara settle on this: “Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task” (2012, p. 197). This definition conflates necessary conditions for crowdsourcing, such as having a task and a crowd to carry it out, with features that are likely to aid the emergence of “wisdom of crowds” effects, such as heterogeneity and varying knowledge. Nonetheless, it suitably frames the general kinds of activities and participants that are involved in crowdsourcing.

While not a necessary feature of crowdsourcing generally, a diverse range of expertise is important in mathematical cases. More diverse crowds will likely also have different approaches to problems, different ideas, and different kinds of expertise. With the potential for insights from different specialisms to mix and contribute, this allows for the “wisdom of crowds” effect to surpass any individual insights mathematicians might have separately. In fact, there is research to suggest that diversity and heterogeneity is more important to overall problem-solving ability than the individual abilities of group members (Hong and Page 2004).

The literature on crowdsourcing spans a huge range of cases and applications, but within mathematics there are obvious candidates, like the *Polymath* projects, *Minipolymaths*, *MathOverflow*, and *Math.Stackexchange*, all discussed in the next section. The broader literature on crowdsourcing usually sees numbers of participants in the thousands or tens of thousands, but in the mathematics cases

the numbers are much smaller: *Polymath1*, for example, had fewer than fifty active participants.

3 The use of crowdsourcing in mathematical research

Mathematics researchers communicate and collaborate online in many ways. Several online tools are central to the practices of research mathematics. *Emails* are the communication method of choice across academia and widely used for communicating written mathematics. Mathematical content is often written in uncompiled TeX markup: for example, Villani's *Birth of a Theorem* (2015) presents many such emails from work that led to his Fields Medal (see also Barany 2010, p. 7). The *arXiv* is the online public repository of choice for many mathematicians, with preprints routinely uploaded prior to official publication. The *arXiv* sends out alerts of new papers to interested colleagues, keeps track of changes, and can help to settle priority disputes. Neither of these is crowdsourcing in itself, although they can provide useful data for researching mathematical practices (for an early study of email use in a formal science, see Merz and Knorr Cetina 1997; for a recent application of corpus-based methods to the *arXiv*, see Mejía-Ramos et al. 2019).

Solutions to (presumptively) *known* mathematics problems can be crowdsourced at question and answer sites. Some of these are geared towards research mathematicians, such as *MathOverflow* and *Math.Stackexchange*. Questions about mathematics can get answers from a userbase which includes many professional mathematicians. Answers are rated by readers, with preferred answers displayed more prominently. The format also encourages discussion on the topic and further questions. So the questioner employs a crowd of research mathematicians to search for a solution to the problem, benefitting from the crowd's expertise and knowledge, and from their collective evaluation of the answers.

Other question and answer forums encompass mathematics at all levels, including undergraduate and earlier mathematics students. Current examples include *thestudentroom*, *reddit*'s mathematics section */r/math*, and many more. Students ask questions, sometimes from curiosity and sometimes for help with coursework. It is not uncommon for talented students to stumble upon an obscure but known insight, and to ask for references to work on the area. These forums thus crowdsource feedback, solutions, and knowledge of the mathematical literature, drawing on the range of expertise in the crowd, the ability of the crowd to effectively find errors in student work, and students can benefit from anonymity to crowdsource their homework answers. These forums also help socialize students to professional norms

and values of the community, as described by Dawkins and Weber (2017). The content of such forums presents a rich and as yet unexplored dataset for mathematics educators.

Crowdsourced mathematics is also more directly applied to collaborative problem-solving: for instance, the *Polymath project*, its spin-off *Mini-polymath*, and the subsequent *Crowdmath* for student work (Gerovitch et al. 2017). The *Polymath* project, initially based at the blog of the well-known mathematician Tim Gowers, used the wisdom of crowds approach to solve high-level, challenging mathematics problems through open, online, massively collaborative work. The first project found an elementary proof of the density Hales–Jewett theorem, leading to a published article (Polymath 2012), and several of the follow up projects have made major headway into open problems. *Mini-polymath* was a way of examining how online mathematics happens, by testing similar methods for simpler problems: participants collaboratively solved International Mathematics Olympiad problems. *Polymath* benefits from crowdsourcing, as the large number of participants and their diverse expertise is vital for some of the leaps of ideas required to solve problems and prove theorems and, at least as importantly, organizing, explaining, and reinterpreting known ideas. In theory, these projects were open to all, but the difficulty of the content skewed the demographics towards professional mathematicians. So, while crowdsourcing is more accessible than a mathematics department breakroom, it can still include or exclude people, and not necessarily based on their mathematical ability (see also Rittberg et al. 2019).

Hence we may perhaps distinguish two varieties of mathematical crowdsourcing: unstructured crowds, where the participants are treated as interchangeable (despite potentially significant variance), and structured crowds, in which there are clear divisions of labour and authority. Q&A sites such as *MathOverflow* are comparatively unstructured but, as we shall see, the *Polymath* projects have much greater internal structure.

4 Quantitative approaches

Quantitative approaches to studying crowdsourced mathematics may be understood in terms of a framework set out by Barany (2009). He suggests that researchers who use numerical techniques to study the nature of knowledge, should follow these steps:

- 1 Capture knowledge in a form that can be analyzed quantitatively,
- 2 Develop means of quantification to match epistemic intuitions, and
- 3 Use mathematical techniques to study these quantifications.

The first step is uncharacteristically easy, since crowdsourced mathematics is usually already in such a form, and, by its nature, is openly available. The third step is likewise straightforward, though involving more work. The difficult step—which can call into question the value of a study—is the second: it faces the problem of construct validity, does the operationalization of a concept actually correspond to that concept? Work in this field can be open to the charge that this step has been reversed; i.e., “develop epistemic intuitions to match means of quantification”. This is an easy trap for the eager scholar hungry for results. Even when sensibly done, a convincing justification as to *why* it is acceptable or useful to represent or measure a given intuition in the stated quantitative way is often missing.

Barany suggests that “such techniques enable one to supplement intuitive or impressionistic analyses by re-framing qualitative problems in quantitative terms” (*ibid.*). This is an important point—the work in itself is not a complete study of some intuition in context, but a *supplement* to qualitative work. Scholars often have the skills to do one type of study and not the other, so we suggest either collaborating or building on existing work. This is made easier—in both the quantitative and the qualitative case—if scholars make public as much of their work as possible, including any data that they have produced.

There are examples of quantitative work that implicitly follow Barany’s guidelines in multiple disciplines, including cognitive history, social psychology, human–computer interaction (HCI), network analysis, and argumentation.

4.1 Cognitive history

Varshney (2012), working within cognitive history seeks to answer the question *Are there differences between individual and collective intelligence?* This speaks to the fundamental question of whether crowdsourced mathematics differs from traditional mathematics, and thereby whether it should be studied as a novel practice or as a proxy for mathematical practice in general. In keeping with the cognitive historical focus on cultural artefacts, he takes five classic theorems³ in geometry from ancient Greek mathematicians as examples of individual intelligence and compares these against the combinatorial proof of the density Hales–Jewett theorem, developed in *Polymath1*, as his example of collective intelligence. We can understand his project in terms of Barany’s framework as follows:

- 1 Represent the individually and collectively produced proofs as argument graphs. In Greek proofs statements are nodes and citations of support between statements are directed edges. In the *Polymath* proof nodes are blog posts and directed edges references among posts.
- 2 Identify epistemic intuitions that may be quantified in terms of argument graphs. Varshney has three:
- 3 Collective intelligence might result in more ideas than individual intelligence; this would show in the structure of the argument graphs, both as multiple ideas supporting a single idea (convergent arguments) and as a single idea supporting a large number of other ideas (divergent arguments).
- 4 Different inference rules will be preferred in each case.
- 5 Greeks will use shorter and simpler arguments.
- 6 Use network science and discrete mathematics to study these quantifications:
- 7 The first intuition can be tested via the degree distributions of the argument graphs. Nodes that are strong hubs (with high in-degree) would correspond to intense integration of information, whereas nodes that are strong authorities (with high out-degree) would correspond to intense dissemination of information.
- 8 The second intuition can be tested by considering the subgraph distributions of the argument graphs.
- 9 The third intuition can also be tested by subgraph analysis, for instance in terms of the number of nodes in a graph.

Varshney confirms his first and second intuitions, but not the third: subgraph analysis shows that *Polymath1* uses short and easy-to-follow arguments to an even greater extent than ancient Greek mathematics.

This is a position paper, and there are limitations of this particular study—some of which are acknowledged by the author.⁴ The methodology and research programme, however, are extremely interesting. Identifying differences between individual and collective intelligence would help us to compare crowdsourced and traditionally produced mathematics. A concrete method of comparing proofs, as represented as argument graphs, within a framework which is sufficiently flexible to allow for other representations and

³ Euclid’s *Elements* II.5, Archimedes’ *The Method* I, Archimedes’ *On Sphere and Cylinder* I.30, Archimedes’ *Spiral Lines* 9, and Apollonius’ *Conics* I.41.

⁴ For instance, in order to focus on differences between individual and collective intelligence, other variables should be eliminated as far as possible. Here, ground-breaking work is compared with everyday work, geometry with combinatorics, a final proof (frontstage) with the discovery of a proof (backstage), modern work with work so ancient we cannot even be sure it was the product of an individual. No justification is given for choosing these particular proofs. Furthermore, the extremely small sample size (one versus five proofs) suffices for a proof of concept at best, not general conclusions (as Varshney sometimes incautiously attempts). It is also, perhaps, unclear how the historical context has been taken into account.

other methodologies used to compare them is a powerful approach.

4.2 Social psychology

Woolley, Chabris, Pentland, Hashmi, and Malone (2010) demonstrate how well-established methods and results in the psychology of individual intelligence can usefully be employed to study collective intelligence. They start from what is “arguably, the most replicated result in all of psychology” (*ibid.* p. 687, see also Deary 2000), that there exists a *general cognitive ability*: people who are good at one cognitive task are likely to also be good at other, unrelated, cognitive tasks. From here, it is relatively straightforward to use the same experimental setup to test whether there exists an analogue general cognitive ability as a feature of a group’s collective intelligence. Via a series of experiments on a total of 192 groups ranging from two to five members, Woolley et al. test their research question: *Do groups, like individuals, have characteristic levels of intelligence, which can be measured and used to predict the groups’ performance on a wide variety of tasks?* They find that group performance over a variety of tasks is a good predictor of performance on a criterion task; while average and maximum intelligence scores of individual group members were not significantly correlated with a general collective intelligence factor c and not predictive of the criterion task performance. They further tested other factors which might be expected to predict group performance, including group cohesion, motivation, satisfaction, social sensitivity, conversational turn-taking and the gender balance in the groups; finding that only the latter three correlate with c .

We can see parallels and differences between this study and our domain of interest. This methodology could be extended to include tests on much larger groups, on “crowds” and on people working together remotely. Both criterion tasks and the initial variety of tasks could be in the mathematics domain. Individual experience in mathematics could be recorded alongside individual intelligence, and so on. Of particular interest is Woolley et al.’s conclusion that “...it would seem to be much easier to raise the intelligence of a group than an individual. Could a group’s collective intelligence be increased by, for example, better electronic collaboration tools?” (Woolley et al. 2010, p. 688). If so, this would present an innovative and cost-effective approach to tough mathematical problems, with important implications for mathematics education, suggesting greater emphasis on group work.

4.3 Human–computer interaction (HCI)

Working in HCI, Cranshaw and Kittur (2011) combine social science theory with in-depth descriptive analysis of data

gathered from *Polymath1*. They use a similar framework to that in Barany (2009) and similar mathematical methods to Varshney (2012) to answer very different research questions: (1) *What factors contributed to the success of Polymath1?* and (2) *How can we design for large scale scientific collaborations?* Since *Polymath1* was a conspicuous success, we should hope to learn from it in the design of future projects.

They get a general feeling of the landscape from simple metrics such as numbers of participants and blog posts per participant, seniority of participants as estimated from Google Scholar citation counts, and participants’ gender. They then consider the role of leadership, coordination, and threading, and the burden to newcomers. This involved defining what these terms mean. In the leadership case, leaders were seen as those who summarised progress, made public judgments as to what was relevant in a previous thread and where they thought the discussion should go next, and made significant contributions to the proof—both in terms of volume and of importance (measured by the quantity of subsequent comments).

In the case of coordination and threading, they considered what worked and did not work about multiple threads and parallel discussions. Parallelism worked on a global level, with work computing exact bounds on Hales–Jewett numbers for small dimensions, led by Tao, going on in parallel with proof attempts of the general theorem, led by Gowers. However, attempts to parallelise work locally on these two projects were largely unsuccessful, with the exception of the interactive “reading group” to discuss background material, which Tao set up. Cranshaw and Kittur further analysed dependencies in the comments, and meta-comments on this aspect, finding that the discussion was largely localized.

On newcomers, they hypothesised that once the project started, the technical nature of the generated content made it difficult for newcomers to join in. They calculated cumulative number of contributors to *Polymath1* over time, finding that the core set started contributing to the project very early. In order to understand the nature of participation and contribution in *Polymath1*, they used the official *Polymath1* timeline, created on the project’s wiki to highlight the comments that were milestones to the proof. Important comments were defined as comments which were milestones or contributed towards a milestone, which triggered lots of other comments, or which linked other contributions in a useful way. They applied a series of graph centrality measures on each node of the comment reference graph. This showed the leadership of Gowers, Tao, and a third potential leader; the existence of participants who made few, but very important comments; and—interestingly—that level of seniority did not correlate to volume or importance of contribution. They conclude with design recommendations for encouraging newcomers, focusing on three challenges: identifying what to read,

identifying tasks to work on, and learning required background material.

This work is a good example of how a framework such as Barany's can be applied sensibly. However, more justification for *why* the measures for concepts such as leadership are appropriate would strengthen this study, as would adding other perspectives on leadership, as one of the authors does in later work. Tausczik, Kittur and Kraut (2014) use three methods to investigate types of collaborative acts, asking *How does collaborative problem solving occur on MathOverflow?* and *What strategies are most successful?* Firstly, they use quantitative analysis, relating collaborative acts to solution quality. Secondly, they apply grounded theory to 150 problems from *MathOverflow* to provide a taxonomy of collaborative acts, coded by whether a contribution provided information, clarified the question, critiqued, revised, or extended an answer. Finally, they conducted semi-structured interviews with 16 active *MathOverflow* contributors to better understand the collaborative acts, the role they played in the collaborations, and how they contributed to the development of solutions. This work shows how combining quantitative and qualitative methods can build up a rich picture of a concept in crowdsourced mathematics.

4.4 Network analysis

Kloumann, Chenhao, Kleinberg and Lee (2016) made a comparative study of the differences between crowdsourced collaboration on open research problems (the *Polymath* projects) and on hard problems with known solutions (the IMO questions on the four *Mini-polymath* projects). These are compared on three axes: the roles and relationships of the authors, the temporal dynamics of how the projects evolved, and the linguistic properties of the discussions. The authors ask the following research questions: *What are the differences between online collaborations on research and on hard problem-solving?*, and *How (and when) can we find whether a comment is a general contribution or a research highlight?* The authors develop a computational model to predict whether a given comment is from an open or difficult but known research problem. It is based on comment length, roles, temporal, and linguistic features, and achieved 90% accuracy. Finally, they consider whether breakthrough comments could be automatically recognised, and if so, how that could be used in real-time to improve the process. One attraction of this approach is its scalability: a computational model could be applied to much larger datasets where hand coding, on which some other approaches depend, would be prohibitively labour-intensive. Real-time identification of breakthrough comments would be of immense value, as

an educational tool and as a facilitator of human–computer collaboration.

4.5 Argumentation

In earlier work (Corneli et al. 2019), we⁵ explored the application of recent argumentation research to crowdsourced mathematics. We asked: *How can we represent mathematical argument using Inference Anchoring Theory (IAT)?*; *How can we extend IAT to give a more complete picture of the linguistic, dialectical, and inferential structure of the arguments?*; and *To what extent can our extended theory (IATC) represent real-world examples of mathematical practice in a way that can make them accessible to computational reasoning?* IAT is designed to model both the inferential structure and the dialogical structure of arguments and how they interact (Budzynska and Reed 2011). We represented a Q&A example from *MathOverflow* and an excerpt from *Mini-polymath1* in IAT, which allowed us to represent dialogue moves, speech acts, and inferences, and gave us a way of connecting arguments to dialogue. However, IAT treats propositions as black boxes, which prevented computational reasoning on mathematical propositions. Hence we extended IAT to IATC, C for content, which allowed us to unpack the propositions, making the relationships between the content of propositions explicit. This yields a more complete picture of the conversation than IAT.

This was a useful exercise to explore the relevance and utility of theories in argumentation which were not developed with mathematical discussion in mind. Although the dataset was small and hand-selected, it felt sufficient to develop and illustrate our new theory, at least as a starting point to answer our research questions.

We also explored the questions *Can we represent Lakatos's informal logic of mathematical discovery as a theoretical dialogue game and then at an abstract level?*; *Can we build a computational model of the theoretical and abstract model?*, and *Does Lakatos's theory apply to crowdsourced mathematics?* (Pease et al. 2017). Our data was the conversation in Lakatos's *Proofs and Refutations* (1976), and hand-picked excerpts from *Mini-polymath3* (2011). We took a practical approach to the first research question, by representing Lakatos's theory as a dialogue and then reasoning over directed graphs. The second research question was largely an engineering problem. In this context it is the third research question that is of the most interest: in some ways the theoretical and computational models developed earlier become tools used for exploring this question. For this question, we hand-coded excerpts from *Mini-polymath3* using

⁵ That is, at least one of us; in this case the first two authors, with other collaborators.

the formalism we developed, to show explicitly how Lakatosian reasoning contributes to the core steps in the development of the proof. This demonstrated that Lakatos-style reasoning can be used to describe at least some real-world examples of crowdsourced mathematical conversations.

Much of our work is motivated by incorporating crowdsourced mathematics into work on social machines, shedding light on how computers can support the process and providing a forum for them to do so; with the ultimate goal of developing new software for human–machine hybrid research teams (Martin and Pease 2013).

5 Qualitative approaches

Crowdsourced mathematics leaves a trail of data which has proved rich enough to warrant qualitative approaches from a number of disciplines, including empirical philosophy, sociology, ethnography, and reflections from mathematicians themselves.

5.1 Empirical philosophy

In Pease et al. (2019) we used close content analysis (Krippendorff 2004) over the comments on the research threads of *Mini-polymath1–4* (2009, 2010, 2011, 2012) to study explanation in mathematics. We started with four research questions (or conjectures), gleaned from the literature on explanation in mathematics: *Is there such a thing as explanation in mathematics?*; *Are all explanations answers to why-questions?*; *Does explanation occur primarily as an appeal to a higher level of generality?*; and *Can explanations be categorized as either trace explanations, strategic explanations or deep explanations?* We supplemented these with two further conjectures which emerged via a pilot analysis of a subset of the data: *Do explanations in mathematics contain purposive elements?* and *Can explanations occur in many mathematical contexts?*

Close content analysis is a qualitative methodology in which presence and meanings of concepts in rich textual data and relationships between them are systematically transformed into a series of results. The method proceeds by analysis design, application, and narration of results. Analyses may be text-driven, content-driven, or method-driven, depending on whether the primary motivation of the analyst is the availability of rich data, known research questions or known analytical procedures. We used keyword indicators to highlight parts of the data to analyse: these were drawn from pre-existing lists of premise, conclusion, and explanation indicators. We performed a complete search of indicators to highlight excerpts which may involve explanation. For each instance with an indicator we used close content analysis,

taking the surrounding context into consideration, to consider each of our six research questions. We then repeated this process on randomly selected parts of the conversation which did not contain indicator terms.

This approach was useful as a way of testing claims made in the mathematical practice literature. It relied on the existence and acceptance of indicator terms; although complementing this with randomly selected examples made the results much stronger. We found evidence for explanation being widespread in mathematical practice, and not just in proofs. This is in apparent tension with the result of (Mejía-Ramos et al. 2019) that explanation is rarer in mathematics than natural science; this may be because they studied a corpus of preprints, whereas our data exhibited crowdsourced mathematics in progress. We also found reasons to doubt some of the most influential philosophical theories of mathematical explanation (and some support for less popular theories).

5.2 Sociology

Barany (2010) asks *In what sense Gowers was really free to make up conventions as he went along?* Looking at the *Polymath1* discussion in depth, Barany considered the narrative of the conversation, the aims of the participants and their goals in re-creating in person mathematical discussion, in the light of the technical constraints and functionalities of the blog medium. In particular, he found that the technology constrained the freedom to make up conventions, but explored the adaptability of the blog platform to the purposes of crowdsourced mathematics.

In Pease and Martin (2012) we used a grounded theory approach to explore the questions *What do mathematicians talk about?* And *How do mathematicians use examples?* Using *Mini-Polymath3* (2011) as data, we found that the mathematical content could be divided into concepts, examples, conjectures, proof, or “other”, with examples forming the biggest single category. A follow-up study used data from *Mini-Polymath1* (2009) to explore how mathematicians use examples. Grounded theory is useful when the research question is very general, open and exploratory, and when the starting point is data, rather than a hypothesis or theory.

5.3 Ethnography

In Lane et al. (2019) and Martin and Pease (2015) we conducted ethnographic studies of crowdsourced mathematics. Ethnographic methods involve close, often immersive, observation of people in their cultural settings. This includes looking at what people say or do not say, how they say it and to whom, what is implicit, how they interact socially and culturally, how meaning is constructed and understood, etc. The outcome of an ethnographic study is a narrative account

of a particular culture within a theoretical context. Given the in-depth nature of ethnographic studies, ethnographers tend not to attempt universal truths, rather focusing on a small number of studies, which they aim to analyse in a detailed and complex way, and in the context of wider cultural and historical factors. Researchers are often immersed in a culture, for an extended period of time, thus participating in the culture themselves. They use observation, interviews and other (usually qualitative) methods to conduct their work.

Ethnographic studies of crowdsourced mathematics are rare. In Lane et al. (2019) we asked *How can we reconcile the contrasting notions of romanticised self-presentations of the isolated genius, with ethnographic studies of mathematicians at work?* As data, we used accounts of Andrew Wiles's proof of Fermat's Last Theorem (Wiles 2000), the *Polymath* projects (the discussion threads, and papers about the project), and placed our observations in a broader literature on landscape, social space, craft and wayfaring, viewing the mathematician in both contexts as crafting a journey through a mathematical landscape. We explored the notion of mathematicians' metaphors of journeys in space and indicated how these might be framed in terms of literary studies, social science, and philosophy, suggesting that ideas of explorations of a fixed landscape might be broadened to consider how mathematicians themselves create that landscape. Theories of craft, in particular Ingold's (2011) notion of crafting as wayfaring, opened up new possibilities for framing the practice of mathematics, shedding further light on the educational role of *Polymath* collaborations.

In Martin and Pease (2015) we asked *How does collaboration enable mathematical advance?*, and *How does crowdsourced collaboration compare to other collaborations in mathematics?* We contrasted *Polymath* with the famous early twentieth-century collaboration between Cambridge mathematicians G. H. Hardy and J. E. Littlewood. As source material we used Hardy's published reflections on his practice (Hardy 1929, 1940), and Littlewood's *Mathematician's Miscellany* (Bollobas 1986), along with personal letters, Hardy's collaborations with Littlewood and Ramanujan (Rice 2015), papers analysing their collaboration (Cartwright 1981; 1985), and research notes between Hardy and Littlewood⁶; and reflective pieces by both *Polymath* participants and "spectators" on the experience of such large collaboration in a public arena, looking at the collaborations, and the institutions and structures that enabled them. Similar themes emerged, such as tolerance of errors, dead ends, and lack of understanding. We argued that the goals of collaborative scholars—emerging in the time of Hardy and Littlewood; established in the time of *Polymath*—are

still a mixture of intellectual satisfaction and professional recognition. Then as now, mathematicians count success as proving significant results and publishing them in significant journals, or the additional recognition of well-known prizes.

Ethnographic studies enable a deep look at a research topic, taking context such as historical and cultural factors into consideration. As with cognitive history, results found using this approach cannot be assumed to generalise but offer an in-depth look at one or two studies.

5.4 Reflections from mathematicians

Many of the crowdsourced mathematics websites contain areas in which the 'crowd' can reflect on the process. Often, in particular in the early days, this was simultaneous with the mathematical collaboration, so that processes could be adapted on the fly, as people explored the new way of working. One key feature of *Polymath* was that it was created by and for mathematicians using a pre-existing technology with which they were already familiar, both in producing and consuming content. The fact that senior figures in the field are prepared to try such a bold experiment, to think through clearly for themselves what the requirements are, and to take a "user centred" view of the design, is striking. For example, Tao's response to the suggestion that participants might use a platform such as *github* (which some argued would simplify the final stage, collaborating on a paper): "One thing I worry about is that if we use any form of technology more complicated than a blog comment box, we might lose some of the participants who might be turned off by the learning curve required."

On the whole they have been successful in solving problems amenable to the approach, with the added benefit of presenting to the public a new way of doing mathematics. The failures have been failures to find the necessary mathematical breakthrough, rather than failures of the *Polymath* format.⁷

The discussion threads of the *Polymath* projects display a record of these discussions, providing a useful resource for those interested in participants' own reflections. These informal discussions are supplemented by published written accounts of the participatory experience (see, for instance, Gowers and Nielsen 2009; Nielsen 2011; *Polymath* 2014).

⁶ Although their rooms at Trinity College were only minutes apart, they would often communicate through notes delivered by a servant.

⁷ Not every crowdsourced mathematics project has been so successful: for example, Gowers's *Tricki*, intended as a collaborative database of mathematical techniques, was outcompeted by *MathOverflow*, contributors being insufficiently motivated without the Q&A format (Gowers 2010).

6 Recommendations

Based on our review of work on crowdsourced mathematics, and our own experiences of studying it, we make the following recommendations (some of broader application):

- 1 **Demonstrate that what is being measured quantitatively is the desired phenomenon.** Scholars should be careful to provide evidence for Step 2 of Barany (2009). Numerical methods are tempting, given their observability, objectivity and so on. However, it is essential to link what these methods measure to what is claimed to be measured.
- 2 **Use multiple approaches to investigate the same phenomenon.** Multiple perspectives on the same research question will yield a deeper, multi-faceted understanding (sometimes called “methodological triangulation”: see Löwe and Van Kerkhove 2019). For instance, the research question addressed in Cranshaw and Kittur (2011): *How does leadership work in mathematical research?* in which leadership is defined and measured in numerical terms, could be supplemented by:
 - 3 Ethnographic studies: for instance, leadership could be seen in the context of metaphors used to describe mathematical activity: journeying, exploring, mountaineering etc., and how leadership works in those domains, with advantages and pitfalls etc.
 - 4 Experiments in social psychology: this could be a comparative study, comparing the leadership functions in crowdsourced and traditional mathematics. It might consider what qualities a leader should have in crowdsourced mathematics, and whether the same people are leaders in traditional as in crowdsourced mathematics. It might also look at whether the concept of leader is stable across a conversation, or whether there are local leaders; varying, for instance, depending on who happens to be online at a particular point, or the current topic of discussion.
 - 5 Interviews: conducting interviews with participants of a crowdsourced conversation, to see how they perceive leadership in that conversation, who the leaders were at various points, looking at whether there is agreement between participants, and so on.
 - 6 A comparative study in which non-crowdsourced collaboration is compared with crowdsourced collaboration; applying the same notions of leadership to both and contrasting how it works in each case.
 - 7 Experimental studies: setting up experiments in crowdsourced collaborations to see, for instance, whether leaders naturally emerge, whether leadership has to be associated with status, knowledge, and so on.
 - 8 Business: identifying different styles of leadership, such as leading from behind, and looking for evidence of those styles in crowdsourced collaborations, testing to see which style is most effective, etc.
 - 9 **Build on other people’s work and make it easy for them to build on yours.** Approaches such as that in (2) will naturally involve experts in different disciplines, so collaboration or building on others’ work is essential. This leads to...
 - 10 **Follow good practices for research into crowdsourcing:**
 - 11 Make data publicly available where possible. Providing annotated data, argument maps, etc. both helps a reader understand prior work and facilitates future work. This must be done with sensitivity: although crowdsourced mathematics conversations are publicly available, there may be information in aggregate which was not apparent from following a conversation. For instance, in a study on errors in crowdsourced mathematics, if it emerged that one participant in particular made many errors, then it may not be politic to publish that study in a way which makes that apparent.
 - 12 Present data in an anonymous form where possible. The public nature of the conversations makes this difficult, but there may be ways to publish a full dataset without it being possible to trace individual participants.
 - 13 As a reviewer, be sympathetic to interdisciplinary approaches. There are well-known difficulties in publishing interdisciplinary work. Of course, there may be valid problems, but reviewers should endeavour to be sympathetic to the goals of the authors.
 - 14 **Minimise confounding variables in comparisons.** To ensure that it is (only) the desired phenomenon that is being compared, minimise variables such as how, why, when, by whom the work was produced.
 - 15 **Justify your selection of data.** There are many instances of crowdsourced mathematics: researchers should state why they selected one dataset and not another (in both comparative and non-comparative approaches).
 - 16 **Look for analogous studies that you can use.** If there are established ways of testing a phenomenon then researchers might find it helpful to adapt these in the case of crowdsourced mathematics, rather than inventing new experimental setups, etc.

7 Issues and limitations of studying crowdsourced mathematics

We should be wary of making generalisations from online to other types of mathematical research.

7.1 Selection bias

To the best of our knowledge no comprehensive, comparative demographic research has been conducted to contrast the mathematicians who take part in crowdsourced mathematics with the larger mathematical population. Nonetheless, we may draw some preliminary observations that suggest these mathematicians may be unrepresentative in several ways.

Firstly, if there is one thing we do know about the mathematicians who collaborate online, it is that they have the time to do so! This may bias the sample towards mathematicians in the early or late stages of their careers: either graduate students and postdocs or full and emeritus professors. Early and mid-career tenure-track professors might be strongly advised to concentrate their efforts on research in which individual contributions are more unambiguously attributable to be sure that they meet criteria for tenure and promotion. While mechanisms may yet evolve to ensure an equitable distribution of credit for the results of massively collaborative research, promotion and tenure criteria notoriously lag behind such innovations.

Secondly, it is reasonable to suppose that not all personality types are equally attracted to online collaborations, especially when such collaborations are essentially public (and, of course, those are the collaborations most accessible to the researcher). In particular, mathematicians who are uncomfortable sharing incomplete ideas may make less effective collaborators: once they have developed an idea sufficiently to be prepared to share it, the opportunity for it to be of use to other collaborators may have passed. If such reticence correlates with other interesting features of mathematical practice, then such features may be absent from studies based upon crowdsourced mathematics. More worryingly, demographic groups that are underrepresented in online spaces or mathematics in general are also likely to be underrepresented in crowdsourced mathematics.

A broader concern with samples of mathematical work drawn from online collaborations is that the participants may be unrepresentative simply by virtue of their ready access to the internet. Researchers in psychology have long complained that the typical samples used in psychological research are drawn exclusively from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies (Arnett 2008). In some respects, internet-based research offers a partial remedy to this problem: online participants recruited through services such as MTurk may well be less WEIRD than undergraduate psychology students, the traditional source of participants for psychological studies (Gosling et al. 2010). Nonetheless, the mathematicians who self-select into online collaboration necessarily have internet access, and presumably fairly frictionless internet access. This may reduce the likelihood of participation in poorer,

less industrialized countries where internet access may be prohibitively expensive and less democratic countries where internet access may be subject to government restriction or international contacts attract official censure.

7.2 Online personalities

One of the earliest and best-known results of research into internet behaviour is the so-called “online disinhibition effect” (Suler 2004). Some people behave in markedly different ways online and offline. Suler, who named the effect, proposed six factors that underpin it: “dissociative anonymity, invisibility, asynchronicity, solipsistic introjection, dissociative imagination, and minimization of authority”. Not all of these factors may be expected to apply to online mathematical collaboration to the same degree as they apply in, for example, social media. Notably, anonymity may have less effect: some collaborations could insist on individuals using names by which they are known professionally (presumably this would be essential if any academic credit is to be assigned) and even where pseudonyms are permitted, they may be expected to be enduring pseudonyms, whose owners have a stake in preserving the reputation associated with the name. Perhaps more importantly still, the best-known mathematical collaborations have minimized status and authority far less than the internet at large. The *Polymath* project, for example, is led by two extremely high-status mathematicians, the Fields Medallists Gowers and Tao, who may be reasonably supposed to carry at least some of their offline authority into their online endeavours. It may be premature to say that the presence of such authority figures is necessary for the success of crowdsourced mathematics, but we know of no examples of such collaborations succeeding without it.

7.3 Limitations

We hope we have shown that crowdsourced mathematics is valuable in the study of multiple aspects of mathematical practice. Nonetheless, there are other aspects where crowdsourced mathematics necessarily provides insufficient data. In particular, we cannot use it to study:

- gestures, intonation, or body language;
- the use of physical materials in mathematical thinking, such as whiteboards, blackboards, and notebooks;
- the use of diagrams, scribbles, or doodles in mathematical thinking.

All of these are important objects of research—and they may all occur *offline* in crowdsourced mathematics too, although not as a means of communication between participants. But crowdsourced mathematics is an increasingly important area of mathematical practice in its own right, with

some notable results to its credit. So, if nothing else, its success demonstrates that not all of these aspects need be shared between participants for mathematics to be done well.

8 Conclusions and further work

Crowdsourced mathematics has been used as an educational resource, as an example of research in action, giving students a chance to look behind the curtains of research—or “see how the sausage is made”, as Tao puts it (quoted in Martin and Pease 2015). At its height, for instance, *Polymath8* was getting three thousand hits a day.⁸ As such, it may affect how mathematics is practiced in future. These experiments may change what comprises mathematics (or indeed, cause it to come full circle, since ancient mathematics was much more like a conversation than mathematics since the invention of the printing press).

The question *Is there any difference between crowdsourced mathematics and traditional mathematics?* is not static and cannot be answered in a binary way. The best we can say is, “in this regard, at this time, given this context, crowdsourced mathematics and traditional mathematics are alike or different”. We hope that this paper provides useful thoughts on how such an answer may be given.

Acknowledgements We are grateful to the anonymous referees, guest editors, and editor-in-chief for their meticulous and invaluable feedback. Ursula Martin and Alison Pease are funded by the UK Engineering and Physical Sciences Research Council under Grants EP/R03169X and EP/P017320 respectively.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

⁸ All the number theory students at a summer school on Bounded Gaps admitted to following *Polymath*. One undergraduate thanked the organisers for such an educational resource and reported “reading the posts and following the ‘leader-board’ felt a lot like an academic spectator sport. It was surreal, a bit like watching a piece of history as it occurred. It made the mathematics feel much more alive and social, rather than just coming from a textbook. I don’t think us undergrads often get the chance to peek behind closed doors and watch professional mathematicians ‘in the wild’ like this, so from a career standpoint, it was illuminating” (Polymath 2014, p. 17).

References

- Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, 63(7), 602–614.
- Ball, P. (2014). Strength in numbers. *Nature*, 506, 422–423.
- Barany, M. (2009). Quantitative epistemology and the structure of logical knowledge. Unpublished. <https://www.mbarany.com/QESLK.pdf>.
- Barany, M. J. (2010). “But this is blog maths and we’re free to make up conventions as we go along”: Polymath1 and the modalities of “massively collaborative mathematics.” *WikiSym '10* (pp. 10:1–10:9). New York: ACM.
- Bollobas, B. (Ed.). (1986). *Littlewood’s miscellany*. Cambridge: Cambridge University Press.
- Budzynska, K., & Reed, C. (2011). Speech acts of argumentation: Inference anchors and peripheral cues in dialogue. *Computational models of natural argument: Papers from the 2011 AAAI workshop* (pp. 3–10). Palo Alto: AAAI Press.
- Cartwright, M. L. (1981). Some Hardy–Littlewood manuscripts. *Bulletin of the London Mathematical Society*, 13, 273–300.
- Cartwright, M. L. (1985). Later Hardy and Littlewood manuscripts. *Bulletin of the London Mathematical Society*, 17, 318–390.
- Corneli, J., Martin, U., Murray-Rust, D., Nesin, G. R., & Pease, A. (2019). Argumentation theory for mathematical argument. *Argumentation*, 33(2), 173–214.
- Cranshaw, J., & Kittur, A. (2011). The Polymath project: Lessons from a successful online collaboration in mathematics. *CHI 2011* (pp. 1865–1874). New York: ACM.
- Dawkins, P. C., & Weber, K. (2017). Values and norms of proof for mathematicians and students. *Educational Studies in Mathematics*, 95(2), 123–142.
- Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain*. New York: Oxford University Press.
- Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2), 189–200.
- Galton, F. (1907). Vox populi. *Nature*, 75, 450–451.
- Gerovitch, S., Braverman, J., & Mirny, A. (2017). Crowdmath: Massive research collaboration among high school and college students. Presented at the Enabling Mathematical Cultures Workshop, University of Oxford, 5th–7th December 2017.
- Gosling, S. D., Sandy, C. J., John, O. P., & Potter, J. (2010). Wired but not WEIRD: The promise of the Internet in reaching more diverse samples. *Behavioral and Brain Sciences*, 33(2–3), 94–95.
- Gowers, T. (2010). Is the Tricky dead? Online at <https://gowers.wordpress.com/2010/09/24/is-the-tricky-dead/>.
- Gowers, T., & Nielsen, M. (2009). Massively collaborative mathematics. *Nature*, 461, 879–881.
- Hardy, G. H. (1929). Mathematical proof. *Mind*, 38, 1–25.
- Hardy, G. H. (1940). *A mathematician’s apology*. Cambridge: Cambridge University Press.
- Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, 101(46), 16385–16389.
- Ingold, T. (2011). *Being alive: Essays on movement, knowledge and description*. London: Taylor & Francis.
- Kloumann, I. M., Tan, C., Kleinberg, J., & Lee, L. (2016). Internet collaboration on extremely difficult problems: Research versus Olympiad questions on the Polymath site. *WWW 2016* (pp. 1283–1292). New York: ACM.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). New York, NY: Sage.
- Lakatos, I. (1976). *Proofs and refutations*. Cambridge: Cambridge University Press.

- Lane, L., Martin, U., Murray-Rust, D., Pease, A., & Tanswell, F. (2019). Journeys in mathematical landscapes: Genius or craft? In G. Hanna, D. Reid, & M. de Villiers (Eds.), *Proof technology in mathematics research and teaching* (pp. 197–212). Cham: Springer.
- Löwe, B., & Van Kerkhove, B. (2019). Methodological triangulation in empirical philosophy (of mathematics). In A. Aberdein & M. Inglis (Eds.), *Advances in experimental philosophy of logic and mathematics* (pp. 15–37). London: Bloomsbury.
- Martin, U., & Pease, A. (2013). Mathematical practice, crowdsourcing, and social machines. In J. Carette, D. Aspinall, C. Lange, P. Sojka, & W. Windsteiger (Eds.), *Intelligent computer mathematics* (pp. 98–119). Berlin: Springer.
- Martin, U., & Pease, A. (2015). Hardy, Littlewood and Polymath. In E. Davis & P. J. Davis (Eds.), *Mathematics, substance and surmise* (pp. 9–23). Cham: Springer.
- Mathoverflow. <https://mathoverflow.net>.
- Mejía-Ramos, J. P., Alcock, L., Lew, K., Rago, P., Sangwin, C., & Inglis, M. (2019). Using corpus linguistics to investigate mathematical explanation. In E. Fischer & M. Curtis (Eds.), *Methodological advances in experimental philosophy* (pp. 239–264). London: Bloomsbury.
- Merz, M., & Knorr Cetina, K. (1997). Deconstruction in a ‘thinking’ science: Theoretical physicists at work. *Social Studies of Science*, 21(1), 73–111.
- Minipolymath1. (2009). IMO 2009 Q6 as a mini-polymath project. Online at <https://terrytao.wordpress.com/2009/07/20/imo-2009-q6-as-a-mini-polymath-project/>.
- Minipolymath2. (2010). Minipolymath2 project: IMO 2010 Q5. Online at <https://polymathprojects.org/2010/07/08/minipolymath2-project-imo-2010-q5/>.
- Minipolymath3. (2011). Minipolymath3 project: 2011 IMO. Online at <https://polymathprojects.org/2011/07/19/minipolymath3-project-2011-imo/>.
- Minipolymath4. (2012). Minipolymath4 project: IMO 2012 Q3. Online at <https://polymathprojects.org/2012/07/12/minipolymath4-project-imo-2012-q3/>.
- Nielsen, M. (2011). *Reinventing discovery: The new era of networked science*. Princeton: Princeton University Press.
- Pease, A., Aberdein, A., & Martin, U. (2019). Explanation in mathematical conversations: An empirical investigation. *Philosophical Transactions of the Royal Society A*, 377, 2140.
- Pease, A., Budzynska, K., Lawrence, J., Reed, C., & Corneli, J. (2017). Lakatos-style collaborative mathematics through dialectical, structured and abstract argumentation. *Artificial Intelligence*, 246, 181–219.
- Pease, A., & Martin, U. (2012). Seventy four minutes of mathematics: An analysis of the third Mini-Polymath project. In A. Pease & B. Larvor (Eds.), *Proceedings of the symposium on mathematical practice and cognition II* (pp. 19–29). Birmingham: AISB.
- Polymath, D. H. J. (2012). A new proof of the density Hales–Jewett theorem. *Annals of Mathematics*, 175(3), 1283–1327.
- Polymath, D. H. J. (2014). The ‘bounded gaps between primes’ polymath project: A retrospective analysis. *Newsletter of the European Mathematical Society*, 94, 13–23.
- Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Murray, P., Schawinski, K., et al. (2010). Galaxy Zoo: Exploring the motivations of citizen science volunteers. *Astronomy Education Review*, 9(1), 1–18.
- Rice, A. (2015). Partnership and partition: A case study of mathematical exchange within the London Mathematical Society. *Philosophia Scientiae*, 19, 5–24.
- Rittberg, C. J., Tanswell, F. S., & Van Bendegem, J. P. (2019). Epistemic injustice in mathematics. *Synthese*. <https://doi.org/10.1007/s11229-018-01981-1>.
- Stillman, G., Brown, J., & Czochoer, J. (2020). Yes, mathematicians do X so students should do X, but it’s not the X you think. *ZDM Mathematics Education*. <https://doi.org/10.1007/s11858-020-01183-5>
- Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & Behavior*, 7(3), 321–326.
- Surowiecki, J. (2004). *The wisdom of crowds*. New York, NY: Doubleday.
- Tausczik, Y. R., Kittur, A., & Kraut, R. E. (2014). Collaborative problem solving: A study of MathOverflow. *CSCW ’14* (pp. 355–367). New York: ACM.
- Varshney, L. R. (2012). Towards a comparative cognitive history: Archimedes and D. H. J. Polymath. In *Collective Intelligence 2012: Proceedings*, arXiv:1204.3534.
- Villani, C. (2015). *Birth of a theorem*. New York: Farrar, Straus & Giroux.
- Wiles, A. (2000). On solving Fermat. <https://www.pbs.org/wgbh/nova/physics/andrew-wiles-fermat.html>. Transcript.
- Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330, 686–688.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.