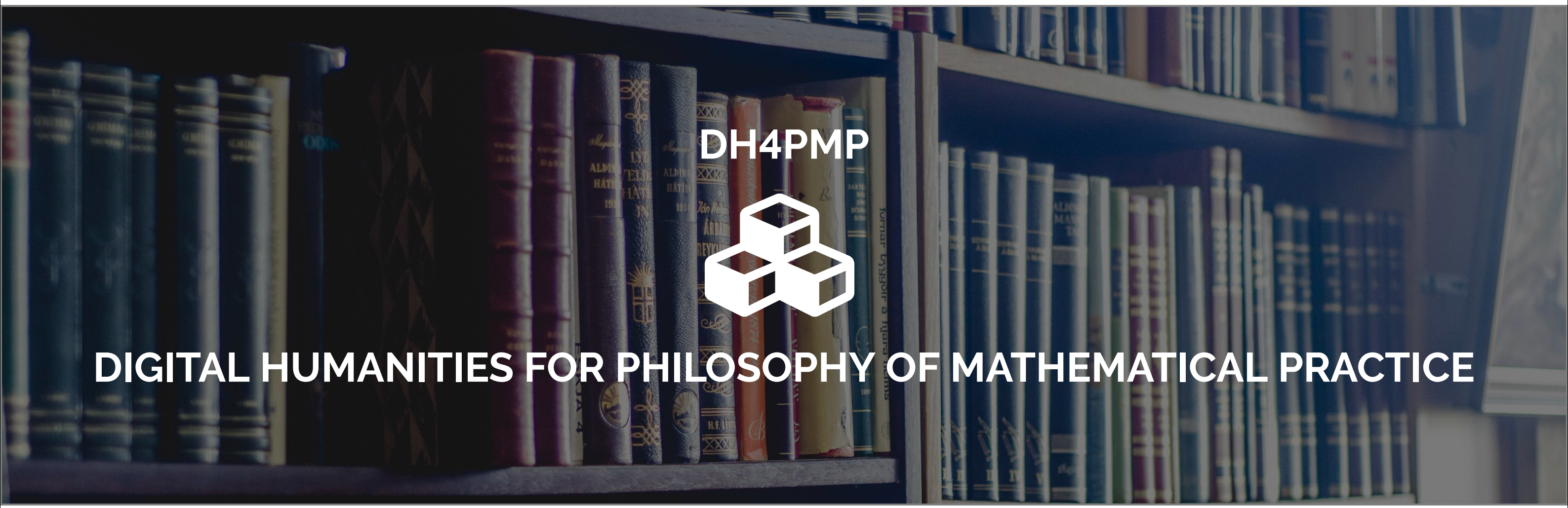

Mining Mathematical Reviews for Empirical Philosophy of Mathematical Practice

Henrik Kragh Sørensen

Digital Humanities for Philosophy of Mathematical Practice, Section for History and Philosophy of Science, Department of Science Education, University of Copenhagen, Denmark. henrik.kragh@ind.ku.dk.

September 2, 2021

Acknowledgements

I am grateful to Mikkel Willum Johansen, Fenner Tanswell, Sebastian Stendal, Chris S. G. Nielsen and Sophie Kjeldbjerg Mathiasen for stimulating discussions during 2020 that led me to write down this note.

1 Introduction

In the field of *Philosophy of Mathematical Practice (PMP)*, we are interested in building philosophical analyses on empirical grounds. This poses a number of methodological questions that the emerging field is only beginning to really grapple with (Aberdein and Inglis, 2019).

Getting access to actual mathematical practice involves a number of methodological problems that we have recently tried to overcome through the use of sociological methods such as interviews and questionaires, anthropological methods such as embeddings and observations, and historical methods such as case studies and oral history.

The increasing digitalization and emergence of *Digital Humanities* as a field has offered new ways to gain empirical access to parts of mathematical practice through tools and methodologies based on digital mining of text and images (see e.g. Jänicke et al., 2015; Terras, Nyhan, and Vanhoutte, 2013), and machine-learning tools such as object detection and natural language processing (Sørensen and Johansen, 2020). However, this approach is still very much in its infancy, and two strands of challenges need to be overcome:

1. We need to overcome technical challenges in processing mathematical products digitally and gain access to collections of mathematical products that lend themselves to *big-data analysis*.

Henrik Kragh Sørensen (Mar. 8, 2021). *Mining Mathematical Reviews for Empirical Philosophy of Mathematical Practice.* DH4PMP Work-in-Progress & Proof-of-Concept Series 10. Version 1. 10 pp. URL: https://www.erda.dk/vgrid/DH4PMP/wippoc/paper10.pdf.

2. And more interestingly and importantly, we need to work on how we may bridge the gap between what can be empirically known and what we are interested in as philosophers of mathematical practice. Thus, it becomes important to explore how we may instrumentalize philosophically research questions to make them accessible through this new set of methods.

As for item 1, published sources may be accessible from scanned or otherwise digitalized traditional journals with a long historical span,[1] or from the recent release of arXiv data onto the machine-learning platform Kaggle.[2]

As for the other item 2, more scholarly work is required, and its form varies across the types of questions and empirical designs that we may consider. An incomplete list of research approaches in this vein would include:

1. Using machine-learning to detect and possibly classify case materials for philosophical analysis (see e.g. Sørensen and Johansen, 2020),

2. Using corpus-linguistics to test hypotheses about e.g. how and how much mathematicians use certain terms (see e.g. Tanswell and Inglis, 2020; Mejia-Ramos et al., 2019), typically used to nuance *a priori* philosophical claims (see also Inglis and Aberdein, 2014),

3. Using text-mining to build categories (e.g. through a form of grounded analysis) of the spectrum of meanings of loaded terms in mathematical texts (see e.g. Dawkins, Inglis, and Wasserman, 2019).

The feasability and applicability of any of these methods to a given research question also hinges critically on the availability of a suitable data set. And typically, we would think of the relevant data set for these kinds of analysis of mathematical practice to be the *primary* mathematical outlets of mathematical knowledge production such as journal articles or research monographs. However, different data, such as textbooks, may also be useful, depending on the research question.

Secondary mathematical outlets such as blogs or online discussion fora have also been analysed and provide important insights into how mathematicians write *about* mathematics and discuss it among themselves in writing (see e.g. Pease, Aberdein, and Martin, 2019; Pease and Martin, 2012). The institutionalized abstracting services of the *Mathematical Reviews (MathSciNet, formerly MR)* and *Zentralblatt Mathematik (zbMath)* offer another fruitful and largely untapped source for enquiry into *secondary* mathematical communication.

2 The data and the metadata

Both MathSciNet and zbMATH are databases set up as *abstracting services* aimed at providing mathematicians with an overview of new publications within their field.[3] The databases thus include bibliographic information such as authors, titles, and publication data etc. for (in principle) *all* mathematical publications, either journal articles or monographs. This information is kept from the inception of the Zentralblatt in 1931 and of the Mathematical Reviews in 1940. Moreover, each service includes a short *synopsis* of the publication, written by another colleague and aiming at describing and (to a lesser extent) evaluating the contents of the publication and its relevance to the field.[4] This synopsis, which is called an *abstract* in the MathSciNet, thus offers a reader's review of the publication at about the time of its production. Finally, each

[1] Such as Göttingen Digitalizierungszentrum (`https://gdz.sub.uni-goettingen.de/`) or JSTOR (`https://www.jstor.org/`) who also offers a *Data for Research (DfR)* initiative.

[2] See `https://www.kaggle.com/Cornell-University/arxiv`.

[3] For the interestingly politicized history of such services, see Siegmund-Schultze (1994).

[4] In some cases of less accessible publications, the synopsis is replaced with the author's abstract.

Henrik Kragh Sørensen (Mar. 8, 2021). *Mining Mathematical Reviews for Empirical Philosophy of Mathematical Practice.* DH4PMP Work-in-Progress & Proof-of-Concept Series 10. Version 1. 10 pp. URL: https://www.erda.dk/vgrid/DH4PMP/wippoc/paper10.pdf.

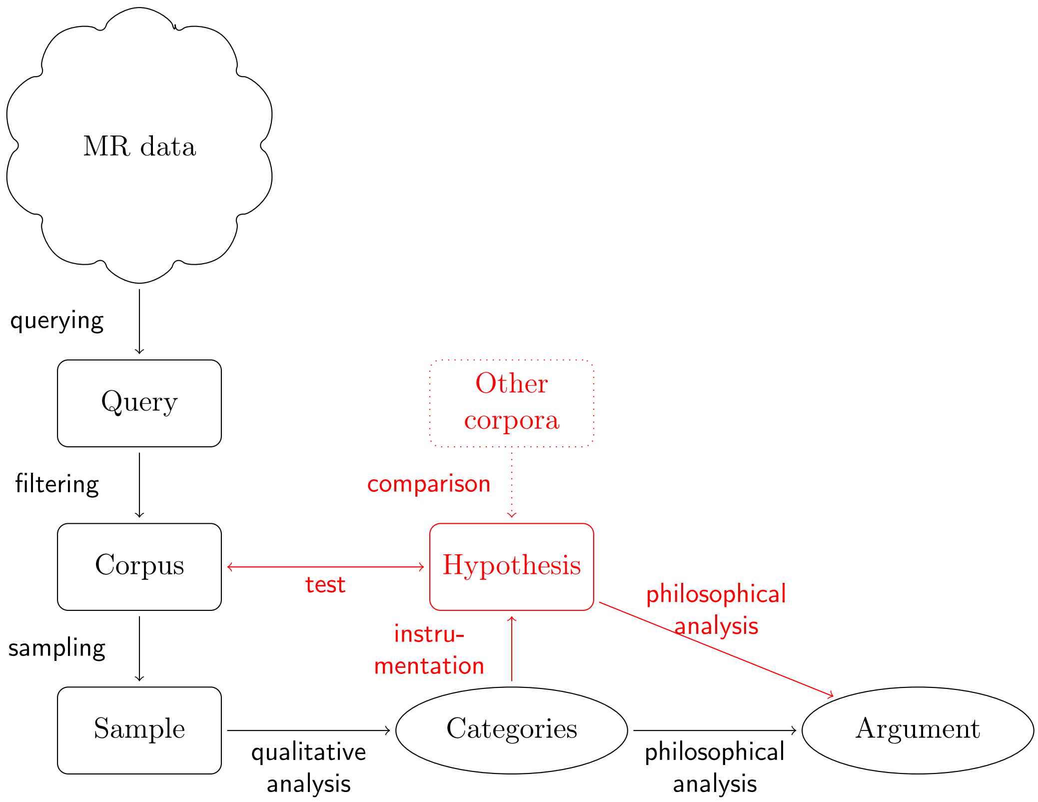

Figure 1: Overview of our research design.

publication is assigned a primary and possibly a number of secondary classifications of the fields of mathematics to which it contributes. These classifications have evolved over time and between the two services, but now a uniform *Mathematical Subject Classification (MSC)* scheme has been adopted and endorsed globally (*MSC2020* 2020).

These abstracting databases thus offer possibilities to ask questions about mathematical reviews and valuations and about the use of certain key terms among mathematical practitioners. Some research have already been done on these databases, but mainly for *scientometric research questions* such as author distribution, collaboration networks and the like (see e.g. Richert, 2011; Brunson et al., 2014). Yet, the method that we are developing aims more at understanding the prevalence and meanings of key elements in the reviews, themselves, interpreted as the primary sources of investigation. Thus, the reviews are our *data* and the bibliographic information about the original publication is to be considered as *metadata* for the present analysis. Obviously, an argument must be made about the relevance and relation of this secondary outlet, which is certainly of sociological and historical interest, to philosophical questions about knowledge production, if we are to take insights from one context into the other.

3 The method

The research design involves managed sampling of reviews, qualitative analysis of a sample, and possibly quantitative hypothesis testing, and is developed for and used on MathSciNet but could easily be used on zbMATH as well. An overview of the process is given in figure 1; the part in red is only relevant, when we want to perform hypothesis-testing, possibly in comparison with other corpora.[5]

[5]For a list of some English corpora, see `https://libguides.bodleian.ox.ac.uk/english-language/Corpora`.

Henrik Kragh Sørensen (Mar. 8, 2021). *Mining Mathematical Reviews for Empirical Philosophy of Mathematical Practice.* DH4PMP Work-in-Progress & Proof-of-Concept Series 10. Version 1. 10 pp. URL: https://www.erda.dk/vgrid/DH4PMP/wippoc/paper10.pdf.

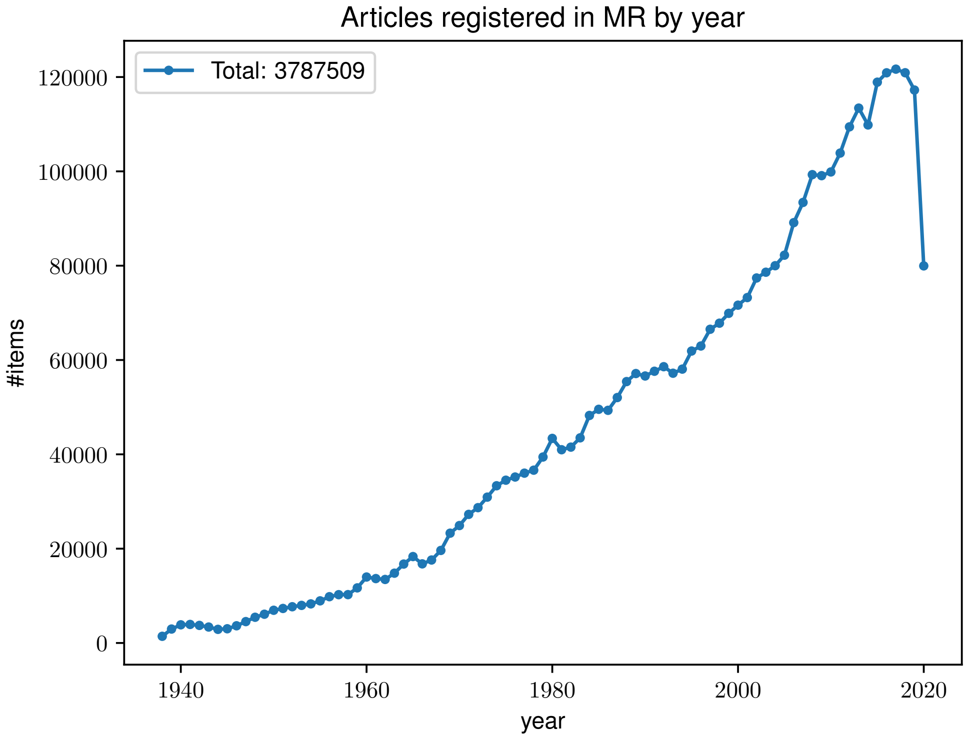

Figure 2: The growth of the mathematical literature indexed in MathSciNet

I. The first step of the process is to *query* the global database with a list of criteria for the abstract to be included in the corpus. These criteria can be formulated in complex ways (see the search form of MathSciNet at `https://mathscinet.ams.org/mathscinet`), combining global queries (*anywhere*) with specialized queries in fields such as author, journal, time span or MSC using logical operations. The purpose of this is to limit the corpus by eliminating entries which are *not* relevant to the research question.

II. The next step is to *visualize* the distribution of the query in suitable ways based on the bibliographic data. Standard visualizations would include distribution over time and over primary MSC code.

III. Based on the visualizations, the query may be *filtered* by more imposing more specific conditions on the bibliographic data, such as eliminating MSC codes or similar. The product of this filtering is the *actual corpus* being studied.

IV. Since the corpus is large, in order to facilitate qualitative analysis, random *sampling* is perfomed. However, the random sampling need not be blind but could strive for e.g. uniformity over time, representation of certain (most frequent) MSC codes or similar. Thus, a *sampling strategy* is required and implemented. The result of this step is the *sample*.

V. The sample can be formatted for easy analysis in software for qualitative analysis such as AntConc (`https://www.laurenceanthony.net/software/antconc/`) for concordancing and text analysis or the general-purpose package NVivo for qualitative analysis and coding.

VI. Based on the qualitative analyses, it may be possible and desirable to formulate a testable hypothesis, e.g. about the concordance of certain words, which can be tested against the

Henrik Kragh Sørensen (Mar. 8, 2021). *Mining Mathematical Reviews for Empirical Philosophy of Mathematical Practice.* DH4PMP Work-in-Progress & Proof-of-Concept Series 10. Version 1. 10 pp. URL: https://www.erda.dk/vgrid/DH4PMP/wippoc/paper10.pdf.

actual corpus (and not just the sample). If the hypothesis involves comparison, say with ordinary language, other corpora may be included.

VII. It is important that all analyses performed in this process are aimed at *philosophical* questions about mathematical practice. This step — which should actually be both first and last — is difficult and challenging as it requires formulating an interesting philosophical research question which is open to this method of enquiry. One possible approach is to take a claim already formulated in the philosophical literature or believed among mathemtical practitiones and test it empirically. Another possibility is to use the process more exploratively to look for nuances in formulation or meaning of key concepts and words. This is likely to be a dialectical proces in which supervision and sparring with colleagues is often most beneficial.

4 A first example: Experiments in the MR

To illustrate the method, I devised an investigation to study the use of experiments in mathematical reviews. The historical and philosophical study of experiments in mathematics is already a flourishing field (see e.g. Starikova and Giaquinto, 2018; Sørensen, 2016; McEvoy, 2013; Arzarello et al., 2012; J. M. Borwein, 2012; Sørensen, 2010; Baker, 2008; Goldstein, 2007; J. Borwein and Bailey, 2004). However, nobody has employed digital humanities to this discussion, and in particular, nobody has looked at the secondary outlets such as MathSciNet, which could potentially help differentiate different meanings and contexts.

To set up the approach, I defined a query to look up the word "experiment" in the "review" field of the MathSciNet database. This query yielded some 33,000 items. The resultant query set is visualized temporally in Figure 3, but given the explosive growth of the mathematics literature (see Figure 2), it may make more sense to compare the corpus to the overall number of items in the MathSciNet database (Figure 4). We can also visualize the distribution of the corpus over different MSC classifications (see Figure 5) where we choose to focus on the primary, main category.

The purpose of the filtering process is to be able to specify conditions that the MathSciNet query format may not support.

To prepare a sample for "close reading" (qualitative analysis), I can randomly sample items from the corpus. The sampling strategy can be e.g. uniform over the entire corpus (i.e. not respecting any difference between the two query terms) or bucket-sampling into e.g. specific MSC categories. When exporting the sample to LaTeX-format, the result is as shown in appendix B.

References

Aberdein, Andrew and Matthew Inglis, eds. (2019). *Advances in Experimental Philosophy of Logic and Mathematics*. Advances in Experimental Philosophy. Bloomsbury.

Arzarello, Ferdinando et al. (2012). "Experimental Approaches to Theoretical Thinking: Artefacts and Proofs". In: *Proof and Proving in Mathematics Education. The 19th ICMI Study*. Ed. by Gila Hanna and Michael de Villiers. New ICMI Study Series 15. Springer. Chap. 5, pp. 97–143.

Baker, Alan (2008). "Experimental Mathematics". *Erkenntnis*, vol. 68, no. 3, pp. 331–344.

Borwein, Jonathan Michael (2012). "Exploratory Experimentation: Digitally-Assisted Discovery and Proof". In: *Proof and Proving in Mathematics Education. The 19th ICMI Study*. Ed. by Gila Hanna and Michael de Villiers. New ICMI Study Series 15. Springer. Chap. 4, pp. 69–96.

Borwein, Jonathan and David Bailey (2004). *Mathematics by Experiment: Plausible Reasoning in the 21st Century*. Natick (MA): A K Peters.

Henrik Kragh Sørensen (Mar. 8, 2021). *Mining Mathematical Reviews for Empirical Philosophy of Mathematical Practice*. DH4PMP Work-in-Progress & Proof-of-Concept Series 10. Version 1. 10 pp. URL: https://www.erda.dk/vgrid/DH4PMP/wippoc/paper10.pdf.

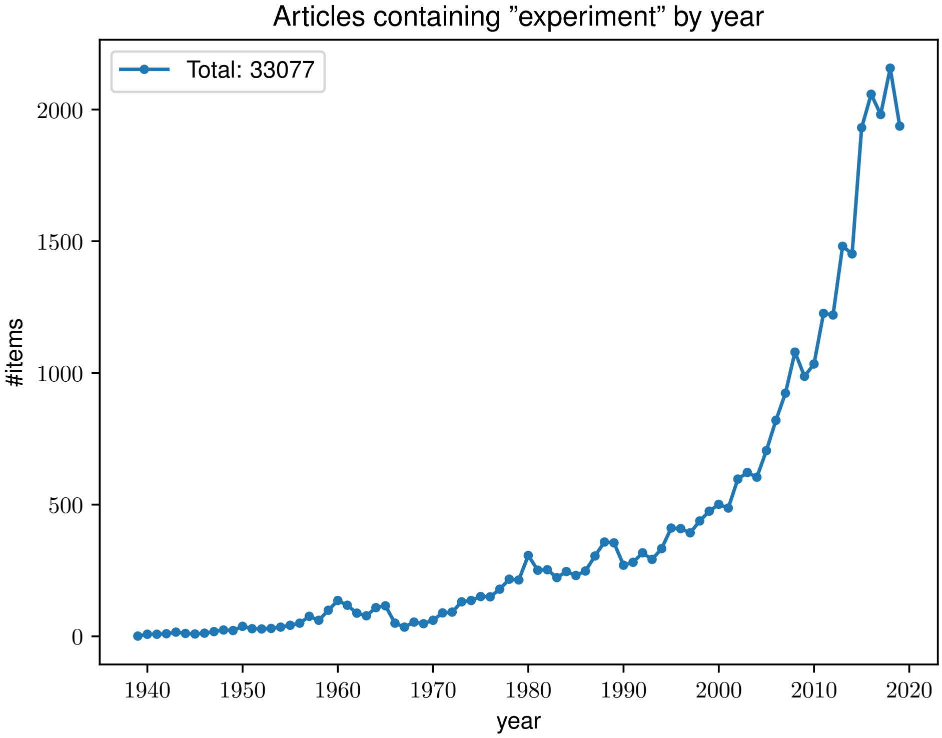

Figure 3: Visualization of the occurance of query terms over time.

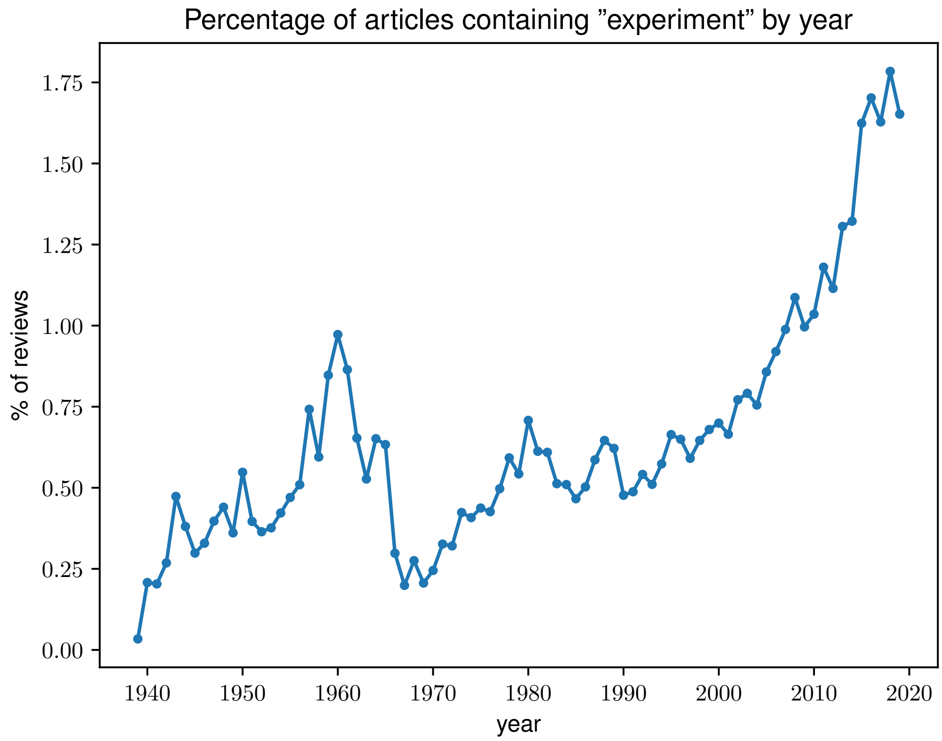

Figure 4: The frequency of search terms normalized.

Henrik Kragh Sørensen (Mar. 8, 2021). *Mining Mathematical Reviews for Empirical Philosophy of Mathematical Practice.* DH4PMP Work-in-Progress & Proof-of-Concept Series 10. Version 1. 10 pp. URL: https://www.erda.dk/vgrid/DH4PMP/wippoc/paper10.pdf.

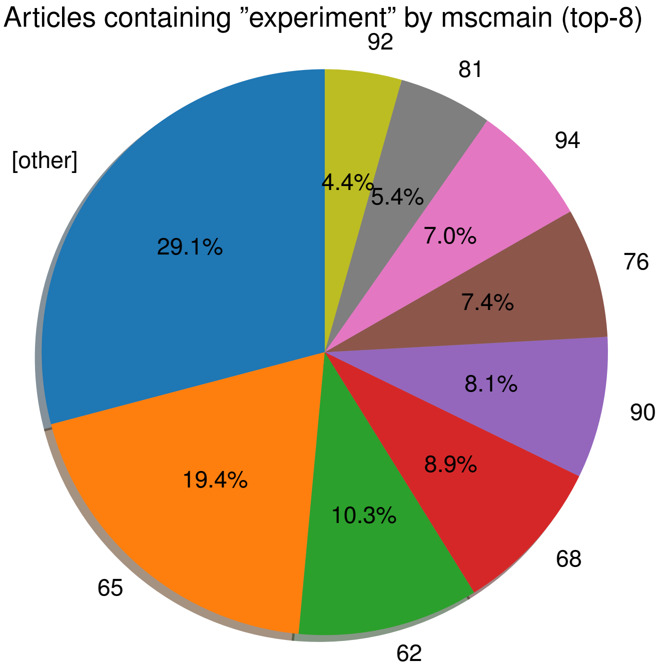

Figure 5: Visualization of the occurance of query terms by MSC code, here only for the top-8 most frequent ones.

Brunson, Jason Cory et al. (2014). "Evolutionary events in a mathematical sciences research collaboration network". *Scientometrics*, vol. 99, pp. 973–998. DOI: `10.1007/s11192-013-1209-z`.

Dawkins, Paul Christian, Matthew Inglis, and Nicholas Wasserman (2019). "The use(s) of *is* in mathematics". *Educational Studies in Mathematics*, vol. 100, no. 2, pp. 117–137. DOI: `10.1007/s10649-018-9868-6`.

Goldstein, Catherine (2007). "How to Generate Mathematical Experimentation, and Does it Provide Mathematical Knowledge". In: *Generating Experimental Knowledge*. Preprint 340. Berlin: Max-Planck-Institut für Wissenschaftsgeschichte, pp. 61–85.

Hanna, Gila and Michael de Villiers, eds. (2012). *Proof and Proving in Mathematics Education. The 19th ICMI Study*. New ICMI Study Series 15. Springer.

Inglis, Matthew and Andrew Aberdein (2014). "Beauty Is Not Simplicity. An Analysis of Mathematicians' Proof Appraisals". *Philosophia Mathematica (3)*, vol. 23, no. 1, pp. 87–109. DOI: `10.1093/philmat/nku014`.

Jänicke, Stefan et al. (2015). "On Close and Distant Reading in Digital Humanities. A Survey and Future Challenges". In: *Eurographics Conference on Visualization (EuroVis). State of The Art Reports (STARs)*. Ed. by Rita Borgo, Fabio Ganovelli, and Ivan Viola, pp. 83–104.

McEvoy, Mark (Feb. 2013). "Experimental mathematics, computers and the a priori". *Synthese*, vol. 190, no. 3, pp. 397–412. DOI: `10.1007/s11229-011-0035-1`.

Mejia-Ramos, Juan Pablo et al. (2019). "Using Corpus Linguistics to Investigate Mathematical Explanation". In: *Methodological Advances in Experimental Philosophy*. Ed. by Eugen Fischer and Mark Curtis. London et al.: Bloomsbury Academic. Chap. 8, pp. 239–264.

MSC2020 (2020). *Mathematics Subject Classification System*. URL: `https://mathscinet.ams.org/msnhtml/msc2020.pdf`.

Pease, Alison, Andrew Aberdein, and Ursula Martin (2019). "Explanation in Mathematical Conversations. An Empirical Investigation". *Philosophical Transactions A*, vol. 377, no. 2140. DOI: 10.1098/rsta.2018.0159.

Pease, Alison and Ursula Martin (2012). "Seventy four minutes of mathematics: An analysis of the third Mini-Polymath project". In: *Proceedings of AISB/IACAP 2012, Symposium on Mathematical Practice and Cognition II*.

Richert, Norman (Aug. 2011). "Authors in the Mathematical Reviews/MathSciNet Database". *Cataloging & Classification Quarterly*, vol. 49, no. 6, pp. 521–527. DOI: 10.1080/01639374.2011.603821.

Siegmund-Schultze, Reinhard (1994). ""Scientific Control" in Mathematical Reviewing and German–U.S.-American Relations between the Two World Wars". *Historia Mathematica*, vol. 21, pp. 306–329.

Sørensen, Henrik Kragh (2010). "Exploratory experimentation in experimental mathematics. A glimpse at the PSLQ algorithm". In: *PhiMSAMP. Philosophy of Mathematics: Sociological Aspects and Mathematical Practice*. Ed. by Benedikt Löwe and Thomas Müller. Texts in Philosophy 11. London: College Publications, pp. 341–360. URL: http://www.lib.uni-bonn.de/PhiMSAMP/Book/.

— (2016). "'The End of Proof'? The integration of different mathematical cultures as experimental mathematics comes of age". In: *Mathematical Cultures. The London Meetings 2012–2014*. Ed. by Brendan Larvor. Trends in the history of science. Birkhäuser, pp. 139–160. DOI: 10.1007/978-3-319-28582-5_9.

— (Mar. 8, 2021). *Mining Mathematical Reviews for Empirical Philosophy of Mathematical Practice*. DH4PMP Work-in-Progress & Proof-of-Concept Series 10. Version 1. 10 pp. URL: https://www.erda.dk/vgrid/DH4PMP/wippoc/paper10.pdf.

Sørensen, Henrik Kragh and Mikkel Willum Johansen (2020). "Counting mathematical diagrams with machine learning". In: *Diagrammatic Representation and Inference. 11th International Conference, Diagrams 2020, Tallinn, Estonia, August 24–28, 2020*. Ed. by Ahti-Veikko Pietarinen et al. Lecture Notes in Computer Science (LNAI) 12169. Springer, pp. 26–33. DOI: 10.1007/978-3-030-54249-8_3.

Starikova, Irina and Marcus Giaquinto (2018). "Thought Experiments in Mathematics". In: *The Routledge Companion to Thought Experiments*. Ed. by Michael T. Stuart, Yiftach Fehige, and James Robert Brown. London and New York: Routledge. Chap. 14, pp. 257–278.

Tanswell, Fenner and Matthew Inglis (2020). "Actions and Imperatives in Mathematical Texts". Manuscript.

Terras, Melissa, Julianne Nyhan, and Edward Vanhoutte, eds. (2013). *Defining Digital Humanities. A Reader*. Farnham and Burlington: Ashgate.

A Specification of the experiment

```
{
    "name" : "sophie",
    "author" : "Henrik Kragh Sørensen",
    "description" : "Experiments in reviews",
    "version" : "1",
    "query" : [
        {
            "description" : "experiment",
             "query" : [ {"review": "experiment*"} ],
             "reviewed" : true
        }
    ],
    "filter" : {
        "method" : "sql",
        "query" : "TRUE"
    },
    "sample" : [
        {
            "id" : "1",
            "method" : "uniform",
            "size" : 100,
            "highlight" :
                {
                    "experiment" : "yellow"
                }
            }
    ]
}
```


Henrik Kragh Sørensen (Mar. 8, 2021). *Mining Mathematical Reviews for Empirical Philosophy of Mathematical Practice.* DH4PMP Work-in-Progress & Proof-of-Concept Series 10. Version 1. 10 pp. URL: https://www.erda.dk/vgrid/DH4PMP/wippoc/paper10.pdf.

B Sample with highlighted terms

Sample extract from corpus experiment

Henrik Kragh Sørensen

March 8, 2021

Contents

0.1 MR2081 (MSC-62, 1940)

It is well known that, in samples of size n from a rectangular population with unknown range θ, the distributions of the range, the mean and largest variate in the sample can be used as a basis for the estimation of θ by intervals. The authors set up a simple sampling experiment, using Tippett's tables of random numbers, for estimating θ from samples of size 2 and 4. 3900 samples of size 2 were drawn and confidence intervals based on sample range, mean and largest variate for a confidence coefficient of .81 were computed. The correspondence between the theoretical and observed results was very close. Similar results were obtained for samples of size 4.

0.2 MR1127 (MSC-70, 1940)

If the Riemannian space-time due to a constant weak gravitational field ϕ is regarded as a flat-space containing a medium whose index of refraction is $1-2\phi$, it is possible to find the motion of the observer relative to the medium by optical experiments. It is suggested that this idea accounts for D. C. Miller's "ether-drift" experiments. The field ϕ is due to the distribution of matter in the universe as a whole.

1

Henrik Kragh Sørensen (Mar. 8, 2021). *Mining Mathematical Reviews for Empirical Philosophy of Mathematical Practice.* DH4PMP Work-in-Progress & Proof-of-Concept Series 10. Version 1. 10 pp. URL: https://www.erda.dk/vgrid/DH4PMP/wippoc/paper10.pdf.