CONSTRUCTED UNDERSTANDINGS OF COLLECTIVE OPINION: INSIGHTS FROM ALGORITHMIC & HUMAN ANALYSIS OF A WIKI SURVEY DATASET

Phoebe Quinn

MEMO TO DISCUSSANT

Thank you very much for taking the time to read this paper. It is a partial, early draft of a journal article (which will also be included as a chapter in my PhD thesis, as I'm doing the thesis with publications format).

I do sincerely apologise that this is not a more complete or polished product – been a real push to get it to this stage in time! I'm hopeful that it is enough of a starting point that you'll be able to see where I'm going.

I thought it might be useful if I provided a bit of context on where this fits within my PhD project below.

I am very much looking forward to hearing your thoughts!

Warm regards,

Phoebe

Context:

- *My PhD research is exploring the use of new forms of civic technology for scaling up deliberation, through action research case studies of the use of the platform Polis for discussion of contentious climate- and disaster-related issues in Australia*
- *My methods for data collection in each case study include:*
    - *A multi-week online Polis conversation*
    - *A background survey when people register for the Polis conversation (demographics etc.)*
    - *Follow up survey(s) for feedback on the process*
    - *Semi-structured interviews and focus groups with selected participants who took part in the Polis conversation and with stakeholders involved in planning, running or using the output*

    - *Ethnographic data through my roles conducting and researching the Polis conversation processes*
- *Analysis includes:*
    - *Analysis of the views expressed during the Polis conversation on the issue at hand*
    - *Analysis of the process of using Polis as part of community decision making, including exploration of:*
        - *how the data was analysed and presented (by the software and by me) i.e. the topic of the below working paper*
        - *the ways in which people participated in the Polis conversation (e.g. participation differences by demographics; how people expressed themselves)*
        - *the ways in which the Polis conversations were planned, conducted and used in the case study settings*
        - *any effects on attitudes, social norms, policy decisions, cohesion, polarisation etc.*
        - *Participants' and stakeholders' subjective experiences of the processes*
- *I am conducting a series of case studies, with insights from the first feeding into the next, and I will be integrating analysis of data across the case studies in later papers*
- *I have conducted one case study so far (a Polis discussion within the University of Melbourne community on the issue 'What should we do about staff air travel emissions?'. I am about to submit my first paper to a journal reporting on the findings.*
- *In February 2024 I will be running the Polis conversation for my next case study, which is in partnership with Monash City Council (a metro LGA in Melbourne's southeast) on the topic: 'As a community, how should we better prepare for and cope with extreme heat?'. This will be open to all who live, study, work etc. in that area.*
- *I will probably conduct a final case study later in 2024 in partnership with East Gippsland Shire Council and the Mallacoota and District Recovery Association, on an issue related to bushfire recovery/preparedness/climate change adaptation.*

INTRODUCTION

Much of the appeal of democratic innovations lies in their potential to more effectively surface consensus on contentious issues, while also deepening understandings of diversity of views (#refs). Consensus and disagreement are *constructed* through democratic processes, not just in the

Commented [PQ1]: Indicates where I'll be adding in references - I've simply run out of time to flesh this part out, unfortunately!

processes through which people develop and express their views, but also in how the opinion data is analysed, synthesised, and presented (#refs). Interrogation of this analytical construction of consensus and disagreement is crucial to democratic studies given the many varied meanings of such concepts and terms (#refs); debates about which of these should be pursued and highlighted (#refs); and role of such decisions in shaping the impacts of democratic processes (#refs).

The increasing use of digital tools and AI in democratic innovations (#refs) both elevates the urgency and broadens the opportunities arising from examination of this analytical construction of consensus and division. The design of such tools requires highly concrete decisions to be made as to how constructs such as consensus and disagreement are to be operationalised within algorithms (#refs). Yet these subjective and consequential decisions are often largely masked in the outputs of automated analysis, as technical appearance tends to generate illusions objectivity (#refs). Unpacking this analytical construction of collective opinion is critical to understanding the implications of using such tools, especially with surging interest in the use of large language models (LLMs) for analysis of opinion data. The crystalised nature of analytical decisions within algorithms also provides fertile ground for theoretical and empirical research around the around consensus and disagreement, generating insights of broader relevance to the study of democratic innovations.

This paper presents insights from an action-research case study of an online discussion in an Australian university, in which the open-source platform Polis was used for collective decision-making on the issue staff air travel and its climate impacts. Polis is a leading example of an emerging class of 'collective intelligence systems' designed to enable mass participation in deliberative democracy through innovations in computing and interface design (C. Small et al., 2021). As a wiki survey (Salganik & Levy, 2015) (a participatory survey in which the items are sourced directly from the collective of participants), it generates a hybrid qualitative-quantitative dataset. Algorithms within the platform instantaneously identify opinion groups and automatically generate outputs synthesising and visualising the opinion data (C. Small et al., 2021).

First, the paper analyses the various ways in which notions of consensus and disagreement were constructed in these automatic outputs, considering what is highlighted and hidden, represented and misrepresented. We then discuss the additional analytical processes conducted by the researchers in an effort to augment the automatic analysis, building upon some elements of the algorithm's output while problematising others. This novel approach to mixed methods analysis of wiki survey datasets thoroughly interrogated the qualitative and quantitative elements of the opinion data, and considered voting patterns on each comment *in context* (of the entire dataset and of the process through which the data was produced). In comparing the insights generated through

the purely algorithmic and the human-algorithmic analytical approach, we conclude by discussing a set of highly pertinent implications for the rapid developments in democratic innovations of our time.

BACKGROUND

CONSTRUCTION OF CONSENSUS AND DISAGREEMENT IN DEMOCRATIC INNOVATIONS

- **Consensus (and disagreement) is always CONSTRUCTED in democratic processes**
    - This applies to:
        - what gets expressed/what data gets produced
        - how the data gets analysed, synthesised and presented (visually and verbally)
        - how the results of the analysis get used and discussed (which I'll also discuss in future papers
    - This is true in analysis of views expressed in processes like mini-publics
        - The extent to which that is obvious and interrogated varies, and there can be more flexibility in it, because it is done more through 'organic'/fluid social processes, rather than pre-determined formulae
    - It is also true in analysis of data generated through tools like Polis:
        - in how these concepts are operationalised behind the scenes in analysis automatically through the platform
        - in additional analyses and interpretations by human analysts
- **Understanding this construction matters ALWAYS in democratic processes, because:**
    - Consensus is often a big selling point for deliberative and democratic innovations, and it is a focus of how the output is used
        - It is a big part of the appeal with mini-publics and climate assemblies (cutting through division/polarisation, reaching negotiated consensus or considered judgement) and with many digital platforms e.g. 'collective intelligence systems' (Ovadya, 2023)
    - There are a wide range of things people mean when they say/hear/think about terms/concepts like 'consensus' and 'disagreement'
        - and there are varying thoughts around which of these matter and should be aimed for in democratic / deliberative processes. Debates around this touch

on: holding some form of consensus (e.g. consensus on a substantive course of action, or meta-consensus) as the goal vs understanding disagreement, pluralism, agonistic democracy etc

- Briefly touch on varied meanings of terms like 'consensus' within the English language (perhaps in comparison with other languages, including how terms like 'rough consensus' has been used in digital democracy in Taiwan to reflect usage within internet communities and more closely approximate a Mandarin term which doesn't not have such a high bar as 'consensus' in English)

- The construction of understandings of consensus/disagreement from opinions expressed substantially affects the output and impacts of democratic processes
  - Looseness around language/concepts of consensus and disagreement leaves the door open for things like:
    - misinterpretation/challenges in communicating about both the processes and the outcomes.
    - people cherry picking in using the output
    - outcomes that don't align with what people (participants and/or other stakeholders) wanted out of the process (e.g. if majority rules consensus is focused on, that can have negative outcomes for minority groups)

POLIS

Polis is an open-source, online platform designed to facilitate large scale deliberative processes (C. Small et al., 2021). It is designed to enable understanding what large groups of people think, in their own words and in real time, enabled by contemporary advances in computing, user interface design and data visualisation. As a data collection tool, Polis is a type of 'wiki survey' – a new class of crowdsourcing tools that source the survey items (i.e. comments) from participants themselves, in real time and in their own words (Salganik & Levy, 2015; C. Small et al., 2021). As such, Polis can be understood as a tool that incorporates some of the features and strengths of participatory, qualitative and quantitative research:

- *Participatory* – in that participants design the survey that they collectively respond to.
- *Quantitative* – in that it enables large groups of people to participate and produces real-time analysis.

- *Qualitative* – in that it enables new ideas to emerge and richer expression than would be possible through a traditional closed question survey approach.

Creators and users of Polis have articulated several deliberate design decisions which have been taken with the intention of fostering pro-social democratic engagement through the ways in which the data are analysed and presented back to participants. For example, the outputs automatically generated include a 'group-informed consensus' metric , which highlights comments that were highly endorsed across all opinion groups, to help to identify 'rough consensus' or promising pathways for practical action, while accounting for minority dissent and protecting against the 'tyranny of the majority' (C. Small et al., 2021). This metric is as a type of 'bridging-based ranking', a set of alternatives to the divisive algorithmic ranking systems which underpin many online platforms (Ovadya, 2022) (with notable recent advances through Twitter's new Birdwatch/Community Notes system which crowdsources factcheck 'notes' to show alongside tweets (Megill et al., 2022; Ovadya, 2022)).

UNDERSTANDING THE ANALYTICAL CONSTRUCTION OF COLLECTIVE OPINION THROUGH DIGITAL TOOLS

- **The construction in the context of tools like Polis warrants particular attention because:**
    - Consensus is certainly big part of the appeal/selling point with Polis - with a focus on mass scale and efficiency
    - It's a relatively new kind of dataset (wiki survey), so analysing/synthesising it requires particular decisions to be made in terms of:
        - computation:
            - specific decisions have (necessarily) been made about how consensus and disagreement should be operationalised in the algorithms (for analysis and visualisation) (in terms of what is accurate/meaningful, and what we want to promote/highlight/hide)
        - language (how the varying numeric instantiations of consensus, disagreement, mixed views are named in algorithmic outputs and how people discus the data)
        - visual representation
    - This very active construction of consensus/disagreement within the software (and the broader way Polis conversations are facilitated etc) is somewhat masked behind an illusion of objectivity (technical-looking outputs on a screen)
    - Emerging research has begun to examine and compare how understandings are constructed through Polis and various approaches to analysing Polis data

        - Refer briefly to **previous approaches** to supplementing the automatic analysis with additional analysis (Moats & Tseng, 2023; C. Small et al., 2021; C. T. Small et al., 2023)
    - Analysing algorithmic analysis can shed light on the construction of consensus in other democratic/social processes
        - This is because of the concreteness of the decisions that need to be made (compared to other processes where construction of consensus happens through, say, in person interactions) - so it's an opportunity to advance understandings / generate insights of broader relevance
    - We are increasingly engaging in the use of algorithms and digital platforms in democratic processes, so understanding what kinds of consensus/disagreement are being forged will matter more and more. This includes:
        - increasing use of Polis and other similar digital platforms
        - increasing interest in integration of LLMs with Polis (and with other approaches to understanding opinions or facilitating democratic discussion) (C. T. Small et al., 2023) i.e. algorithmic analysis not increasingly extending to analysis of qualitative opinion data
    - This paper provides a set of highly pertinent insights to inform consideration of such approaches, through comparison of outputs from a purely quantitative algorithmic analysis and hybrid human-algorithmic, qualitative-quantitative in-depth analysis.

METHODS

GENERATION OF THE POLIS CONVERSATION DATASET

A Polis conversation was held at the University of Melbourne for a two week period in September 2022, in which 173 participants (of whom 142 were staff and the remainder were students) discussed the issue 'What should we do about staff air travel emissions at the University of Melbourne?"

Over a two week period, the participants submitted 239 comments and 22,897 votes. Participants expressed their views by 'voting' (selecting 'Agree', 'Disagree', or Pass/Unsure') in response to comments other people had made on the staff air travel issue, and optionally submitting their own comment(s). The comments were displayed anonymously, popping up one by one. Participants could vote on as many or as few comments as they wished and were encouraged to participate for a

few minutes at a time on multiple occasions so they could engage with new comments as the conversation evolved. Commenting was disabled two days before the close of voting, to enable voting to consolidate around existing comments.

The webpage included a visualisation of the 'opinion landscape' and a link to an interactive report displaying data across all comments, both of which were generated automatically within Polis and updated in real time as the conversation evolved (see section 4.4.1).

Comments were moderated by the lead researcher (PQ) in consultation with authors LG, KB and CG where required, according to a moderation protocol designed to protect against inappropriate comments (e.g. containing hate speech or profanities) and improve flow (e.g. to split comments containing multiple ideas). Comments 1 to 9 were submitted by the researchers as 'seed comments' expressing a range of possible perspectives to get the conversation started.

Commented [PQ2]: Note to discussant: this purple text is self-plagiarised from my other draft paper (Appendix C), just for now in order to ensure you have the necessary info

The lead researcher (PQ) was primarily responsible for facilitation and moderation of the Polis conversation with input as required from co-researchers LG, NC, CG and KB, as well as stakeholders within the University such as the Sustainability Strategy Team.

Through this process, a wiki survey dataset was generated, with a structure known as a 'sparse opinion matrix': it is an *opinion matrix* because there is a data point for the opinion expressed by each participant for each comment ('agree', 'disagree', 'pass/uncertain' or no vote), and it is *sparse* because it is characterised by a high degree of missing data across the matrix since the platform shows comments to each participant one by one in a semi-random order, and participants do not need to respond to each comment.

The study was approved by the University of Melbourne Human Ethics Committee (project ID 23743).

ANALYSIS

STAGE 1: ALGORITHMIC ANALYSIS

The Polis platform automatically analyses the dataset as it is generated, as detailed by Small and colleagues (2021). The machine learning algorithms use Principal Component Analysis (PCA) and k-means clustering to identify 2-5 opinion groups – clusters of people who tend to vote similarly to each other on the various comments – and analyse vote frequencies on each comment according to these opinion groupings. Although the wiki survey dataset contains qualitative comment text data as well as quantitative vote data, the automatic analysis consists of purely *quantitative* analysis of the

structured ‘sparse opinion matrix’ dataset: the algorithms are agnostic to the content of comment, and there is no form of natural language processing (NLP) (although possible future integrations of large language models (LLMs) within the platform are being explored (C. T. Small et al., 2023)).

The results of these algorithmic analyses are presented in two automatic outputs: a detailed automatic report (‘automatic report’) which displays the vote data on all comments across the conversation (Fig. XX), along with the results of the opinion group clustering, through a series of interactive visualisations (Appendix A); and a summary visualisation (Appendix B). Both automatic outputs can optionally be displayed to participants during the Polis conversation, as this occurred in this case study. The algorithmic analyses within Polis also affect the order in which comments are displayed to participants.

STAGE 2: HUMAN-ALGORITHMIC, QUALITATIVE-QUANTITATIVE INTEGRATION

With the aim of informing next steps taken within the University on the issue of staff air travel in this action-research process, we began by consulting the outputs of the Polis automatic analysis. However, for reasons illustrated in the results section of this paper, we decided to conduct substantial additional in-depth thematic analysis to generate the insights required in this action research case study. In the absence of previously established methods for in-depth, holistic interpretation of the qualitative (comment text) and quantitative elements (agree, disagree or pass/unsure votes frequencies) of a wiki survey dataset, we developed a novel approach as discussed below. Visualisations of the qualitative and quantitative data for each comment (Fix XX) were produced to support this analysis[1]. This analysis was primarily undertaken by the lead author whose multiple roles and identities of relevance – as Polis conversation facilitator, researcher, staff member and graduate student at the University of Melbourne – created a unique positionality which played an important role in the construction of interpretive insights from the data. The results of this analytical process discussed in this paper are reported in Quinn et al. (Appendix C).

Commented [PQ3]: Note: this is the first paper from my PhD which I’ll be submitting to a journal very shortly.

[1] [1] We imported these into the software application Obsidian (v1.1.9-1.3) using its mind-mapping and bidirectional linking features to assist in the remaining analytical steps.

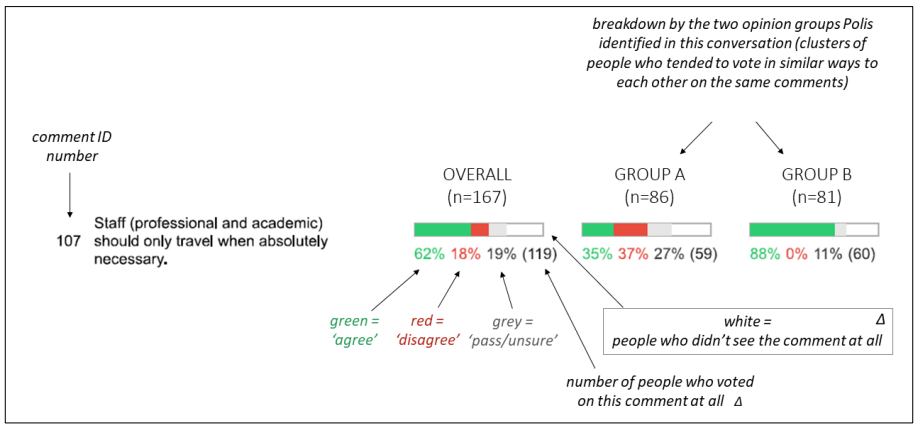


Figure XX – Annotated display of votes data (for comment 107 as an example) explaining how to interpret the figures featured in section 1.2. Δ Response numbers and white space varies across the comments because Polis shows comments to each participant one by one in a semi-random order, and people did not need to respond to each comment.

COMPARISON OF INSIGHTS FROM THE TWO ANALYTICAL APPROACHES

Having completed the interpretive analysis and reported the findings (in a report circulated within the University and a paper submitted for peer review – see Appendix C), we conducted a comparative analysis of the ways in which understandings of collective opinion were constructed and reflected in the outputs from the purely algorithmic and the human-algorithmic analysis. This included revisiting the detailed report and summary visualisation automatically produced by the Polis algorithms and considering the inferences these algorithmic outputs foster. We then explored the similarities and differences compared to the inferences drawn out in the outputs generated through our additional human-algorithmic, qualitative-quantitative thematic analysis of the same dataset.

COMPARATIVE INSIGHTS FROM THE TWO STAGES OF ANALYSIS

ALGORITHMIC CONSTRUCTIONS OF COLLECTIVE OPINION

The automatic report included a list of all the comments in the Polis conversation, with accompanying bar charts depicting the number and percentages of agree, disagree and pass votes and the degree of missing data. In this sense, it provided quite a comprehensive dataset from which one could approach meaning-making with a high degree of autonomy. Nonetheless, there were several important ways in which algorithmic analysis behind the automatic Polis outputs conveyed and encouraged particular understandings of collective opinion in the conversation, foregrounding certain constructions of consensus and disagreement.

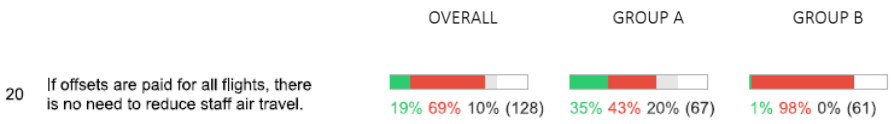
First, constructions of consensus and disagreement in the automatic outputs were based on the treatment of each comment as an atomised, independent unit of analysis, with no algorithmic engagement with the qualitative comment text. The outputs therefore featured individual comments in isolation from other comments, with no links drawn between that touch on similar themes or use similar or different rhetorical approaches.

A core feature of both the summary visualisation and the automatic report was the portrayal of participants as clustered into two distinct opinion groups (Group A and B). It is notable – yet *not noted* in in the Polis outputs – that the algorithm is set to always identify between two and five opinion groups in any conversation: an example of a highly concrete and consequential decision in the algorithmic construction of division and consensus. The summary visualisation featured a graph of the two opinion groups on an abstract 'opinion landscape', each of which could be clicked on to reveal data on the comments (and associated vote frequencies) that characterised that opinion group. In the automatic report, extended versions of each of those features also appeared, and additionally the bar charts visualising the vote frequencies for each comment in the comment were broken down into votes from Group A and B members, as well as overall amongst the whole participant cohort (Fig XX). These bar charts provided scope for a consumer of the report to inspection of patterns of opinion both *between* and *within* the algorithmically-identified opinion groups, if desired. Nonetheless, the above-mentioned structural features of the automated outputs had the effect of highlighting the existence of the two opinion groups and encouraging attention to similarities and differences *between* them first and foremost.

Another way in which the automatic Polis outputs fostered particular understandings of collective opinion was through metrics used to highlight particular comments and determine the order in which comments were displayed in lists in the report. We identified at least 8 different metrics used across the various sections of the automatic report to highlight or order comments according to frequencies of agree, disagree or pass votes. For example, a dot plot at the top of the automatic report presented comments on a spectrum from 'Consensus statements' to 'Divisive statements'; a 'Majority' section features (some) comments with high disagreement or agreement overall on a majoritarian basis (i.e. agnostic to opinion grouping); and the list of all statements in the report can be ranked by the 'group-informed consensus' mentioned earlier. These metrics represent only a tiny proportion of the many conceivable ways that this data could be sliced, especially considering permutations based on the opinion group clustering and missing data.

There were varying (but always quite low) levels of transparency to the casual report consumer in regards to the various metrics used throughout the algorithmic outputs, and often required

substantial effort on the part of our research team to investigate. The challenges in automatic analysis of this data were evident in several quirks in the report from this conversation, whereby certain comments were highlighted in ways that were clearly inconsistent with the intention of the report section. For example, the following comment appeared as the second most 'divisive' on the dot plot in the detailed report:



Yet this same comment also appears in the 'Majority' section designed to highlight 'what most people agreed with' according to a threshold that 60% of participants overall either agreed or disagreed (as explained in the report) and what appears to be additional hidden algorithmic criteria which prioritise a balance of comments with majority agreement and majority disagreement in the set of 10 comments selected to feature here.

HUMAN-ALGORITHMIC, QUALITATIVE-QUANTITATIVE CONSTRUCTIONS OF COLLECTIVE OPINION

We now discuss the ways in which our additional analysis yielded different constructions of collective opinion compared to that fostered by the automatic outputs alone. This is attributable to the more holistic approach taken through our interpretive integration the qualitative and quantitative elements of the data, drawing upon both human and algorithmic insights.

In our additional analysis, we carried forward the algorithm's identification of the two opinion groups, which we saw as a useful starting point for understanding the diversity of views in the conversation. We broadly characterised the differences in opinion between the groups as follows: Group A members tended to express more concern or resistance regarding possible changes or restrictions to reduce staff air travel, while Group B was more strongly in favour of a wide range of actions to reduce staff air travel emissions.

However, in the interest of exploring the contours of agreement and disagreement both *within and between* these groups, we chose to minimise the emphasis on these two groups in the structure of our analysis, eschewing, for example, the analytical possibility of developing more in-depth opinion profiles of these two groups.

Instead, we organised our analysis according to subtopics and themes within the Polis conversation, highlighting matters on which rough consensus was found and areas where views were mixed within

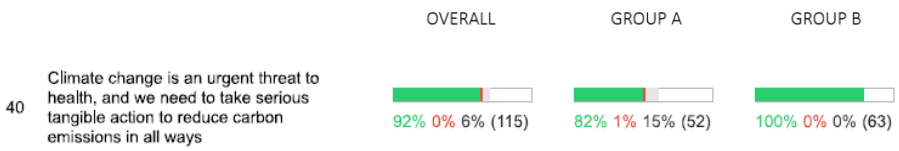
each subtopic. By considering comment and vote data based on subtopics and themes, we were able to consider possible interpretations of vote patterns based on a wide range of factors including the substantive meaning of comments, nuances in phrasing, and possible contextual influences outside the comment itself (e.g. consideration of the possible influence of the role the real-time participants were seeing may have played, and other factors in the broader case study context). This had manifold benefits, as we will see, including minimising the risk of misinterpretation of the data, enhancing confidence in interpretations through triangulation, and surfacing themes and principles with explanatory power and which would not have been evident based on cursory inspection of the data.

Further, in contrast to the rigid metrics and thresholds for consensus and division instantiated within the Polis algorithms, our additional analysis applied an interpretive approach to the vote frequency data: we holistically considered the various elements of the data including rates of agreement, disagreement, uncertainty and missing data figures (for the overall participant cohort and specifically for opinion group A and B). This enabled far greater flexibility to consider the many varied forms and extents of consensus, disagreement and uncertainty, and to construct a narrative designed to surface insights of most relevance to decision-making in the action research case study setting.

We now illustrate these forms of additional insight generated through the human-algorithmic, qualitative-quantitative analysis, through paradigmatic examples.

SKETCHING CONTOURS OF OPINION BETWEEN THE EXTREMES OF CONSENSUS AND DIVISION

Many of the comments highlighted in the automatic report as points of consensus (in the dot plot, the 'Majority' section or the group-informed consensus list) boldly advocated for urgent action to reduce carbon emissions across all sources, e.g.:



A cursory inspection of consensus as constructed in the automatic outputs might therefore foster the impression of strong support for emissions reduction measures, including flight reduction, in both opinion groups. Yet the automatic report highlights points of division that might leave one with

a very different impression: that of a highly divided community in which strong action was widely supported amongst Group B and mostly rejected by Group A. For example:

		OVERALL	GROUP A	GROUP B
66	Stringent criteria should be applied for University support of air travel - see it as a privilege used only in exceptional circumstances	 40% 38% 21% (123)	 6% 65% 27% (58)	 70% 13% 15% (65)

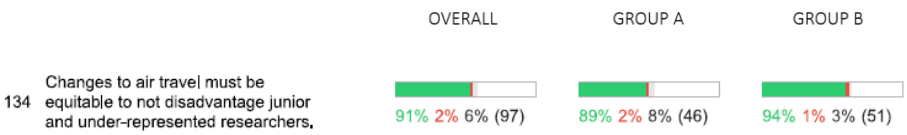
Our in-depth analysis delved into the contours of opinion in the vast middle area between the extremes of consensus and division highlighted in the automatic report, sketching out inferences to make meaning out of data points such as this which may initially appear incongruous. This generated several forms of important insight which can dramatically shape the inferences and actions likely to flow from collective discussions such as this.

First, we found that on many of the most widely endorsed comments in the conversation, the appearance of consensus was largely misleading: it could be attributed in large part to ambiguity or vague framing, rather than reflecting a more robust, genuine form of consensus on a mutually-understood course of action. For example, apparent consensus on sweeping statements like comment 40 above dissipated whenever comments made more specific assertions regarding implications for reducing air travel, where responsibility lies (e.g. at individual or institutional levels), or the relationship between air travel emissions and other emission reduction avenues.

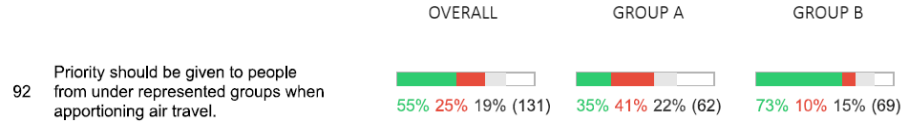
Similarly, among comments relating to possible mechanisms for enforcing reductions to staff air travel, those that were the most broadly endorsed were the most ambiguous, using vague phrasing regarding whether or not ultimate decision-making should remain with the individual (c237 and c61). Through holistic analysis of vote data across comments on this topic, we found that that the extent of disagreement from Group A was linked to the degree of rigidity and infringement on individual autonomy implied.

The above example also illustrates another form of important insight from our analysis: the identification of underlying ideas or principles with explanatory power in interpreting opinion patterns across various comments. In this case, attitude to autonomy was one such principle that emerged as significant in shaping the conversation, showing up implicitly more often than explicitly in comments. Such principles can be understood as interpretive contours in the metaphorical opinion landscape, helping us to chart the space between extreme consensus and extreme division.

The importance of attitudes toward individual versus institutional control was further evidenced through analysis of comments relating to equity in how travel opportunities should be spread amongst staff. In the automatic report, comment 134 was the fourth-highest ranked comment on the group-informed consensus metric:

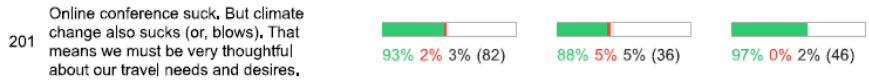


It is easy to imagine a scenario in which the data on this comment alone may be interpreted as reflecting broad support for a wide range of equity-focused initiatives which prioritise travel opportunities for junior staff. However, an important nuance was revealed in comparing votes data all comments on this matter of equity: comments advocating that inequities should not be *exacerbated* were broadly endorsed (c134, c58), whereas comments proposing *active prioritisation* on equity grounds (c1, c92, c95) were much more contentious, drawing significant levels of disagreement especially from Group A, e.g.:

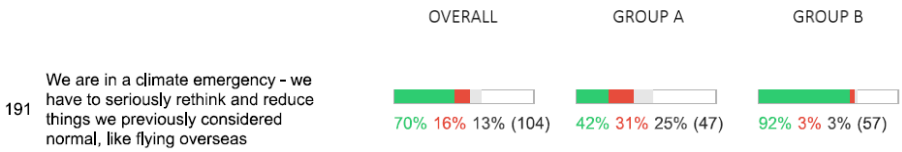


This demonstrates the value of holistic analysis *across topics* in the conversation, as well as amongst comments within each topic, to enable exploration of potential interpretive value of principles such as individual autonomy and institutional control.

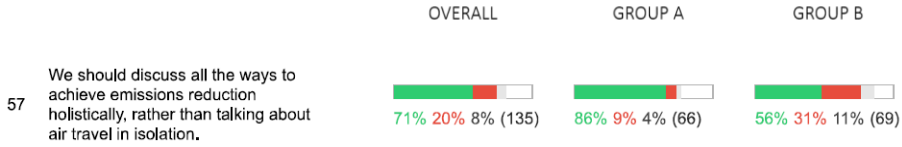
Beyond the evidence that ambiguity was often linked with broad consensus, our analysis also revealed other instances in which discursive features of comments appeared to explain their popularity or unpopularity. Again, analysis of thematically similar comments alongside each other was found to protect against misinterpretation likely to arise from the atomised, low-context treatment of individual comments in the algorithmic outputs. This is of critical importance given the inherent characteristics of wiki survey data sets: since comments are sourced from participants in their own words, there is high variability in phrasing (in stark contrast to, say, validated survey measures). For example, consider comment 201, which was the top comment according to the group-informed consensus metric:



However, viewing the data on related comments together, we noticed that the popularity of this comment was an anomaly amongst comments advocating for flight reductions specifically (c191, c202, c3, c107, c167, c178, c265, c210, c259, c103), and concluded that this was probably attributable to its humorous or gentle phrasing (being 'thoughtful'), in contrast to others which expressed a similar idea in more forthright terms, e.g.:



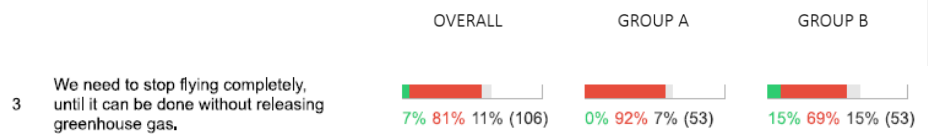
In interpretating the data, we also considered that unlike the context-agnostic algorithmic analysis of the data, participants must naturally have been making contextual inferences when casting their votes on each comment. For example, based on feedback data (to be reported in future publications), we were conscious of the possibility that phenomena such as groupthink and polarisation may have been shaping the data generated, in part fostered amongst some participants by features of the real-time summaries being reflected back to them. Evidence of this was seen most strongly in responses to the following comment:



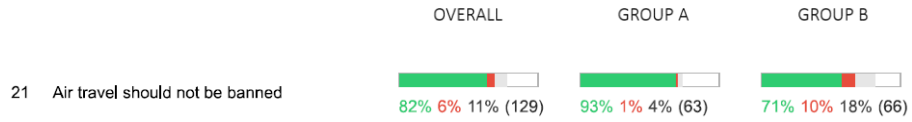
On face value, the substantial rate of disagreement amongst Group B appears to contradict this group's general pro-environmental stance. However, taken in the context of broader disagreements in the conversation (around the relationship between reductions to flights versus other emissions sources, and especially Group B's overwhelming agreement with criticisms of perceived resistance to flight reductions (c210 and 265)) we inferred it was probable that some members of Group B had interpreted comment 57 as an effort to stall change on air travel, and were objecting to that more so than to the comment's explicit meaning.

Finally, our interpretive, holistic approach to the votes data made space for attention to minority dissent, and the flexibility of our narrative approach to presenting the findings enabled nuanced

articulation of the various forms and degrees of disagreement, division and mixed views that fell between pure consensus and extreme division. In this, we also considered the qualitative data and the relative importance of minority dissent on a given comment. For example, comment 3 was highlighted in the 'Majority' section of the automatic report:



Given the relatively radical stance reflected in this comment, we saw it as notable that 30% of Group B members who voted on it agreed or were uncertain (and therefore conveyed this in the outputs of our analysis), especially since those patterns are consistent with a similar but inversely-framed comment:



This illustrates the ways in which comparison of vote patterns across comments was used to evaluate inference, as well as the flexibility of this approach to exploring and conveying nuances in collective opinion across the Polis conversation data.

LIMITATIONS, IMPLICATIONS ETC

TBD

Will include:

- Elaboration of the links between the forms of construction of consensus, disagreement etc with conceptualisations and debates in the literature
- Discuss how our analytical approach differs from those previously used with Polis data (mentioned in Background)
- Discuss implications for consideration of use of LLMs for automated analyses of opinion data expressed through online tools such as this, including that:
    - our findings illuminate the value of integration of qualitative data into analyses

- our analysis process was resource intensive so the possibility of partial automation to make this more feasible holds appeal
- but our findings illustrate the analytical complexities (including the role of context, and the influence of my subjective interpretations through my positionality as researcher, Polis conversation facilitator, and staff member and student at the university) – they shed light on some of the features and criteria we may focus on in experimenting with and appraising the analytical performance if processing down this path

CONCLUSION

TBD

APPENDICES

- Appendix A will consist of a link to the automatic report and a static list of all the comments (including the visuals of the associated vote data) in order or comment ID
- Appendix B will consist of a link to the closed conversation webpage (which shows the 'simpler' opinion map that was embedded under the comment box)
- Appendix C – near-final paper draft reporting the results of my analysis of the opinions expressed in the Polis conversation (attached)

REFERENCES

Carr, H., Smith, J., & O'Brien, T. (2020). *Polis and the Political Process.* Demos. https://demos.co.uk/project/polis-and-the-political-process/

de Fine Licht, K., & de Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making. *AI & SOCIETY*, *35*(4), 917–926. https://doi.org/10.1007/s00146-020-00960-w

Duberry, J. (2022). Artificial intelligence and democracy: Risks and promises of AI-mediated citizen–government relations. In *Artificial Intelligence and Democracy*. Edward Elgar Publishing.

Megill, C., Barry, E., & Small, C. (2022). *"Coherent Mode" for the World's Public Square* (arXiv:2211.12571). arXiv. https://doi.org/10.48550/arXiv.2211.12571

Moats, D., & Tseng, Y.-S. (2023). Sorting a public? Using quali-quantitative methods to interrogate the role of algorithms in digital democracy platforms. *Information, Communication & Society*, 1–35.

Ovadya, A. (2022). *Bridging-Based Ranking: How Platform Recommendation Systems Might Reduce Division and Strengthen Democracy*. Belfer Center for Science and International Affairs, Harvard Kennedy School.

Ovadya, A. (2023). 'Generative CI'through Collective Response Systems. *ArXiv Preprint ArXiv:2302.00672*.

Salganik, M. J., & Levy, K. E. C. (2015). Wiki Surveys: Open and Quantifiable Social Data Collection. *PLOS ONE*, *10*(5), e0123483. https://doi.org/10.1371/journal.pone.0123483

Small, C., Bjorkegren, M., Erkkilä, T., Shaw, L., & Megill, C. (2021). Polis: Scaling deliberation by mapping high dimensional opinion spaces. *Recerca. Revista de Pensament i Anàlisi*, *26*(2), 1–26. http://dx.doi.org/10.6035/recerca.5516

Small, C. T., Vendrov, I., Durmus, E., Homaei, H., Barry, E., Cornebise, J., Suzman, T., Ganguli, D., & Megill, C. (2023). Opportunities and Risks of LLMs for Scalable Deliberation with Polis. *ArXiv Preprint ArXiv:2306.11932*.