Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Narrowing and Recovery of Aspect-sentiment Diversity in Online Discussions

Xiaodan Wang

xiaodan.wang@autuni.ac.nz

Auckland University of Technology

Yanbin Liu

University of Technology Sydney

Jianing Xia

Universal Higher Education, Australia

Yining Wang

University of Melbourne

Minh Nguyen

Auckland University of Technology

Weihua Li

Auckland University of Technology

Shiqing Wu

City University of Macau

Quan Bai

University of Tasmania

Research Article

Keywords:

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11486267/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Narrowing and Recovery of Aspect-sentiment Diversity in Online Discussions

Xiaodan Wang[1,2], Yanbin Liu[3], Jianing Xia[4], Yining Wang[5], Minh Nguyen[2], Weihua Li[2*], Shiqing Wu[6*], Quan Bai[7]

[1]School of Engineering, Yanbian University, Yanji, Jilin Province, China.
[2]Department of Data Science and Artificial Intelligence, Auckland University of Technology, Auckland, 1010, New Zealand.
[3]Australian Artificial Intelligence Institute, University of Technology Sydney, Sydney, 2007, Australia.
[4]Universal Higher Education, Melbourne, Australia.
[5]Faculty of Science, University of Melbourne, Melbourne, Australia.
[6]Faculty of Data Science, City University of Macau, Macau SAR, China.
[7]School of Information and Communication Technology, University of Tasmania, Hobart, 7001, Tasmania, Australia.

*Corresponding author(s). E-mail(s): weihua.li@aut.ac.nz; sqwu@cityu.edu.mo;

Abstract

Online discussions can keep growing while the range of evaluations they express contracts. We study this *tunnel vision*, the progressive concentration of a discussion on a smaller set of aspect-sentiment positions, and ask whether narrowing is transient or lasting and whether receding positions return. We develop a framework that links frozen, thread-specific aspect schemas and aspect-conditioned sentiment labels to whole-thread diversity trends, local contraction episodes and the identities of the positions expressed, and apply it to 2,967 event-centred Reddit threads from April 2019 drawn from a processed corpus of 6,128 threads. Observed tunnel vision occurs in 45.9% of threads, but only 9.3% narrow more than expected under a null that preserves labelled mention density. Local contractions are common (73.8% of threads) and usually reversible (83.7% of first episodes recover), whereas this density-robust minority recovers less often (60.9% versus 86.6%) and, after recovering, more often contracts again. Onset-aligned trajectories show

that the groups diverge after onset rather than in their initial decline. Recovery is accompanied by the return of previously expressed positions and a more even distribution among continued positions, with relatively small changes in the share of first-observed pairs. Richer openings are associated with higher final diversity despite slightly steeper subsequent declines, separating relative narrowing from a narrow endpoint. An independent April 2018 sample reproduces these patterns. The results establish recovery as a central dimension of conversational diversity and provide empirical targets for explaining how online discussions sustain evaluative breadth.

Keywords: computational social science, online discussions, tunnel vision, aspect-based sentiment analysis, temporal dynamics

1 Introduction

Online discussions allow people to share and evaluate different aspects of events. However, continued participation does not necessarily maintain a diverse range of evaluations. As discussions unfold, attention may become concentrated on fewer evaluative positions, even as new comments continue to appear. This narrowing may persist or reverse over time, raising questions about how discussions lose and regain evaluative diversity [1].

Previous research has examined concentration in online discourse through echo chambers [2], cross-platform interaction patterns [3], collective attention [4, 5], and semantic repetition [6]. While these studies offer insights into how information and attention become concentrated, less is known about how evaluative diversity changes within individual discussions. In particular, when does narrowing occur, how often does diversity recover, and which evaluative positions contribute to that recovery?

We examine these dynamics in terms of *tunnel vision*, the progressive narrowing of a discussion toward a smaller set of aspect-sentiment pairs [7]. Each pair combines a specific event aspect with the sentiment expressed toward it, representing a distinct *evaluative position*. For example, support for and criticism of a policy's effectiveness constitute different positions, as do evaluations of its effectiveness and its fairness. By

capturing both what is evaluated and how it is evaluated, this representation allows us to examine diversity beyond topical coverage or sentiment alone. Importantly, tunnel vision here describes the distribution of expressed positions within a discussion, not individual participants' cognitive state. A discussion may therefore exhibit tunnel vision even when participants continue to disagree or use varied language, as long as their evaluations become concentrated on fewer positions.

However, an overall decline in evaluative diversity does not fully capture how a discussion evolves. Discussions with similar downward trends may follow different paths: some narrow early and remain concentrated, others narrow later, while still others experience repeated cycles of contraction and recovery. Moreover, a declining trend does not necessarily mean a discussion ends with low diversity, as those that start with greater diversity may remain relatively broad despite narrowing. Understanding recovery requires looking beyond changes in overall diversity. Diversity may increase because new evaluative positions emerge, previously expressed positions return, or existing positions become more evenly represented. These changes reflect different ways in which a discussion regains evaluative breadth. Distinguishing them requires tracking not only how diversity changes over time, but also which positions are expressed and how their relative frequencies evolve.

To examine these dynamics, we analyse event-centred Reddit threads using frozen, thread-specific aspect schemas and aspect-conditioned sentiment classification. This approach provides a consistent representation of evaluative positions within each thread, allowing us to measure whole-thread diversity trends, identify local contraction and recovery episodes, and examine how individual positions contribute to renewed diversity. We also introduce two null comparisons to assess whether observed narrowing holds up under alternative explanations based on labelled mention density and comment order. Among 2,967 threads in the April 2019 analysis sample, 45.9% exhibit observed tunnel vision, but only 9.3% narrow more than expected under a null model that

preserves mention density. Local contractions are common but usually temporary. Threads exhibiting density-robust tunnel vision recover less frequently and are more likely to contract again following recovery. When recovery occurs, it is accompanied primarily by the reappearance of previously expressed positions and a more balanced distribution among continuing positions, rather than by newly observed positions. We further find that discussions beginning with greater evaluative diversity tend to narrow slightly faster but remain more diverse at their endpoints. An independent sample from April 2018 reproduces these main patterns.

The contributions of this paper are summarised as follows.

1. **A framework for measuring within-discussion tunnel vision.** We combine event-specific aspect-sentiment representations with whole-thread diversity trends and explicit null comparisons, distinguishing observed narrowing from narrowing that remains robust to specified alternative explanations.
2. **A temporal analysis of contraction and recovery.** We distinguish overall tunnel vision from local contraction episodes, showing that temporary narrowing is widespread, whereas density-robust tunnel vision is less common and associated with less frequent recovery and more frequent renewed contraction. We also observe these patterns in an independent sample.
3. **A characterisation of how evaluative diversity recovers.** By tracking the identities and relative prominence of evaluative positions, we show that recovery primarily involves the return of previously expressed positions and rebalancing among continuing positions, with comparatively limited contributions from newly observed positions. This analysis also clarifies why discussions can narrow over time without necessarily ending with low diversity.

The rest of this paper is organised as follows. Section 2 reviews related work on information diversity, attention dynamics and evaluative representations. Section 3 describes the data, measurement framework and analytical methods. Section 4 presents

the findings on whole-thread narrowing, contraction and recovery dynamics, changes in evaluative positions, and the role of opening diversity, together with robustness checks and replication results. Finally, Section 5 discusses the implications, limitations and conclusions of the study.

2 Related Work

This section reviews four strands of research relevant to tunnel vision in online discussions, including information diversity and ideological concentration, attention and discussion dynamics, the representation of evaluative positions, and the distinction between descriptive patterns and their explanations. These studies provide the conceptual and methodological foundations for examining how evaluative diversity narrows and recovers within evolving discussions.

Information diversity and ideological concentration

Research on echo chambers, filter bubbles, and polarisation examines how users, information sources, and viewpoints become separated in online environments [2, 8–10]. Network-based studies investigate how information circulates within like-minded groups and how controversy develops between them [11–13]. Experimental evidence further suggests that exposure to opposing views can sometimes increase rather than reduce polarisation [14]. These studies reveal how online environments shape exposure to different perspectives. However, exposure to diverse viewpoints does not necessarily mean that a discussion represents a diverse range of evaluations. For example, participants may strongly disagree on one aspect of an event while overlooking others, or discuss multiple aspects without forming opposing groups. Tunnel vision captures this distinction by focusing on the distribution of expressed aspect-sentiment pairs within a discussion, complementing existing research on population- and network-level concentration [1]. Previous work introduced the concept of tunnel vision and examined

it in controlled settings [7]. This study extends that work by investigating how tunnel vision develops and changes over time in real-world online discussions.

Attention and discussion dynamics

Research on collective attention shows that public interest shifts across topics over time [4]. At the discussion level, studies examine how conversations grow, how participation unfolds, and how replies are structured [15, 16]. Other research identifies recurring interaction patterns across platforms [3] and shows that highly engaging events can change both the pace and semantic repetition of online discourse [6]. These existing studies reveal how participation and discussion content evolve over time. However, changes in activity or language do not necessarily indicate whether a discussion maintains a diverse range of evaluations. Our study addresses this distinction by tracking aspect-sentiment pairs over time, examining both overall trends in evaluative diversity and local patterns of contraction and recovery.

Representing evaluative positions

Topic models identify semantic themes in text through word co-occurrence or contextual embeddings [17, 18], while aspect-based sentiment analysis identifies the sentiment expressed toward specific aspects [19–21]. The latter is particularly suitable for studying tunnel vision, as a single comment may express different sentiments toward multiple aspects of an event. However, the aspects discussed vary across events, making a fixed set of predefined categories difficult to apply universally. Large language models offer a flexible approach to identifying event-specific aspects and are increasingly used for text classification and coding in computational social science [22, 23]. Nevertheless, their outputs depend on task design, and their agreement with human judgment varies across tasks [24, 25]. To keep temporal measurements consistent, we separate aspect schema generation from sentiment classification and keep each thread's schema fixed throughout the analysis.

From observed trajectories to explanations

Social feedback can reinforce certain forms of online expression [26], potentially contributing to the repetition of particular evaluations. However, repetition alone does not reveal whether alternative positions disappear permanently or reappear later in a discussion. Before explaining why these changes occur, it is important to establish how evaluative diversity changes over time. Describing observed patterns is a distinct scientific objective from predicting outcomes or identifying causal mechanisms [27]. Our study therefore focuses on describing how evaluative diversity narrows and recovers over time, without identifying the causes of these changes or making assumptions about participants' underlying beliefs or emotions.

3 Data and Methods

We analyse individual Reddit threads in chronological comment order. Each aspect-sentiment pair represents an evaluative position; its relative frequency measures expressed positions rather than exposure or beliefs. Supplementary Methods S1–S8 provide detailed definitions, implementation, sample provenance and human validation.

3.1 Corpus, study unit and representation

The data come from the Pushshift Reddit archive for April 2019 [28]. The full corpus consists of 6,128 collected and processed Reddit threads, comprising 3,821 census threads with archived schemas and annotations and 2,307 additional threads sampled from other length bands within the same communities. Of the full corpus, 4,219 threads contain at least 150 usable comments and constitute the reliably measurable corpus. The primary analyses, however, focus on the census frame rather than the entire reliably measurable corpus.

Within the 3,821 census threads, 2,981 meet the minimum requirement of 150 usable comments, providing sufficient windows for temporal analysis. Fourteen of these

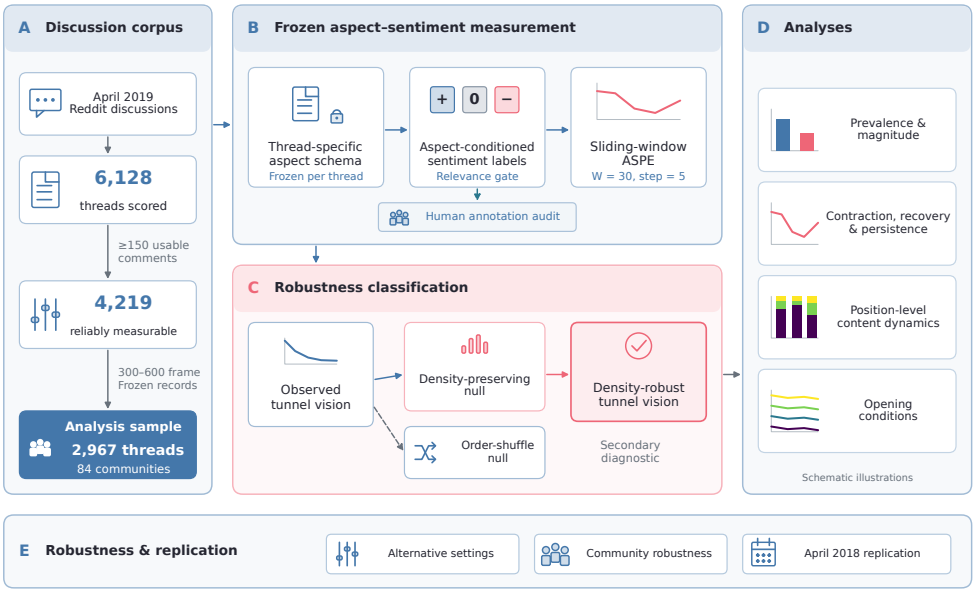


Fig. 1 Study design. (A) Corpus: 6,128 scored April 2019 threads, 4,219 with at least 150 usable comments, and the 2,967-thread analysis sample from 84 communities in the 300–600-comment frame. (B) Measurement: a frozen per-thread aspect schema, aspect-conditioned sentiment labels behind a relevance gate and subject to a human annotation audit, and 30-comment sliding-window ASPE advanced in steps of five. (C) Robustness classification: observed tunnel vision is tested against the density-preserving null, yielding density-robust tunnel vision; the order-shuffle null serves as a secondary diagnostic. (D) The four analyses of Sections 4.1–4.4. (E) Sensitivity checks, community robustness and the April 2018 replication.

threads contain 602–3,239 usable comments despite having platform-reported counts no greater than 600. Because the reported counts reflect snapshots taken at archiving, these threads fall outside the intended usable-comment range and are set aside as supplementary long threads, with no inference drawn from them. The final analysis sample therefore comprises 2,967 threads from 84 communities, containing 1,133,533 usable comments.

3.2 Aspect-sentiment measurement

Figure 1 summarises the study design. Each thread's aspect schema and comment labels are produced once and reused unchanged to calculate windowed diversity and concentration.

Event-adaptive aspect schema

We use a two-pass procedure with `gpt-5-mini` to construct an event-specific aspect schema for each thread. The first pass identifies relevant aspects and provides their names, stance-neutral definitions and lexical cues, while the second pass reviews and refines the schema before it is frozen (Supplementary Methods S7). Because the schema is constructed using comments sampled from across the entire thread, the resulting measurement is retrospective, and entropy is interpreted within each thread's fixed schema rather than on a common semantic scale across threads. Within-thread ASPE rank trends are unaffected by positive rescaling of entropy at the thread level. Empirically, schema size shows only weak associations with these trends and retained equivalent diversity (Supplementary Results R3).

Aspect-conditioned sentiment and pair stream

Each comment is evaluated against its thread's frozen schema using a DeBERTa-v3 aspect-based sentiment classifier [29–31]. A cue-relevance gate selects comment–aspect matches, and each retained match contributes one positive, neutral or negative (a, s) pair. A comment may contribute several pairs and express different sentiments toward different aspects, and neutral labels remain distinct positions. Supplementary Methods S7 specifies the gate, classifier and process checks.

In an independent audit of 501 comments from 42 threads, labelled by two annotators (Supplementary Methods S8 and Results R2), agreement is assessed at three levels: whether a comment addresses an aspect (detection), which polarity it expresses, and whether agreement changes with position in the thread. The annotators agree on whether an aspect is addressed in 91% of comment-aspect cells ($\kappa = 0.48$) and, where both record a pair, on its polarity in 74% of cases. The labelling pipeline (schema, gate and classifier) is evaluated on the consensus cells, those to which the annotators assigned the same label; disputed cells are excluded from the pipeline metrics rather than counted as errors. For detection, the pipeline identifies 78% of the 216 consensus

pairs at aspect level (detection recall), and 56% of its own detections in those cells are consensus pairs (detection precision). For polarity, it assigns the same polarity as the annotators to 96% of the consensus pairs it detects, so that 75% of consensus pairs are reproduced with both aspect and polarity (exact-pair recall). For position, the audit found no evidence that detection declined from the early to the late third of a thread: detection recall is 0.74, 0.79 and 0.81 and detection precision 0.55, 0.51 and 0.62 in the early, middle and late thirds, with early-minus-late differences of -0.08 and -0.07 whose intervals include zero, and exact-pair recall follows the same pattern (0.68, 0.78 and 0.79). Inter-annotator agreement serves as a reference benchmark, and separate comparisons with each annotator are provided in the supplement.

Sliding-window diversity and concentration

Temporal measurement uses a fixed window of $W = 30$ consecutive comments, advanced in steps of five. This window length follows the earlier tunnel-vision formulation and validation [7]. It is treated as part of the measurement definition and is not varied in this study. For window j of thread i, let $q_{ij}(p)$ be the share of retained pair mentions assigned to position $p = (a, s)$. Windows with no retained mention are treated as missing. The core diversity measure is aspect-sentiment pair entropy (ASPE),

$$H_{ij} = -\sum_{p} q_{ij}(p) \log_2 q_{ij}(p), \tag{1}$$

measured in bits: H_{ij} is zero when every mention in the window concerns a single position and increases as mentions spread more evenly over more positions. It is reported with equivalent diversity $D_{ij} = 2^{H_{ij}}$, the number of equally weighted positions that would yield the same entropy, following the effective-number interpretation of diversity [32]. We additionally calculate $C_{ij}^{(5)}$, the share of window mentions held by its five most frequent pairs. Full window indexing and the empirical pair distribution are given in Supplementary Methods S1.

For each thread, the whole-thread diversity and concentration trends are

$$\rho_{\mathrm{ASPE},i} = \mathrm{Spearman}(j, H_{ij}), \qquad \rho_{\mathrm{conc},i} = \mathrm{Spearman}(j, C_{ij}^{(5)}). \tag{2}$$

Each trend lies between -1 and 1 and records only the direction and consistency of change over the thread, not its size: negative $\rho_{\mathrm{ASPE},i}$ indicates that windowed diversity tends to fall as the thread proceeds, whereas positive $\rho_{\mathrm{conc},i}$ indicates that the leading positions tend to take a growing share.

A thread shows *observed tunnel vision* when it meets the pre-specified operational criterion $\rho_{\mathrm{ASPE},i} < -0.15$ and $\rho_{\mathrm{conc},i} > 0$. The -0.15 cutoff is inherited from the earlier corpus-screening protocol and was not selected from the results reported here. Threads that do not satisfy this criterion are *non-tunnel-vision threads.*

Trend and magnitude are kept separate. If $H_{i,\mathrm{early}}$ and $H_{i,\mathrm{late}}$ are the first and last valid full-window entropies, then

$$\Delta H_i = H_{i,\mathrm{late}} - H_{i,\mathrm{early}}, \qquad R_{D,i} = 2^{\Delta H_i} = \frac{D_{i,\mathrm{late}}}{D_{i,\mathrm{early}}}. \tag{3}$$

Thus $R_{D,i} = 1$ denotes no net change, $R_{D,i} < 1$ a narrower final window, and $R_{D,i} > 1$ a more diverse final window.

3.3 Null comparisons and robust classifications

Null comparisons assess robustness to specific alternatives, conditional on the features each randomisation preserves [33].

The *density-preserving null* addresses the possibility that ASPE declines because later comments contain fewer identifiable pair mentions. It preserves the number of pair mentions at each comment position and the overall frequency of each pair within the thread, while randomly redistributing pair labels across the fixed mention

slots. The *order-shuffle null* tests whether the observed trend could arise from the random ordering of comments. It randomly shuffles comment order while keeping each comment's complete set of pair labels intact, including comments without any retained labels. The exact randomisation procedures are detailed in Supplementary Methods S2.

Each null uses 300 fixed-seed permutations and a one-sided empirical probability (Supplementary Methods S2). An observed-tunnel-vision thread has *density-robust tunnel vision* when $p_i^{\text{den}} < 0.05$, *order-robust tunnel vision* when $p_i^{\text{ord}} < 0.05$, and *dual-null-robust tunnel vision* when both conditions hold. These robustness descriptors are not mutually exclusive, and a thread retains its observed tunnel vision classification even if it does not meet either null-specific criterion.

The permutation probabilities indicate how unusual each observed trend is under the corresponding null model, conditional on the features it preserves. They assess robustness to specific alternatives rather than the probability that tunnel vision is present or absent, and do not establish causal mechanisms. We focus on density-robust tunnel vision because it provides stronger evidence of narrowing beyond the specified mention-density alternative. The comparison group, *all other threads*, includes non-tunnel-vision threads and observed-tunnel-vision threads that do not meet the density-preserving criterion. These groups describe different whole-thread trajectories, and their recovery contrasts do not establish causal effects or eliminate measurement error.

3.4 Local contraction, recovery and persistence

Whole-thread trends describe overall changes in evaluative diversity but do not reveal when contraction begins or whether diversity subsequently recovers. Local dynamics are therefore measured by the relative diversity $r_{ij} = D_{ij}/D_{i,\text{ref}}$, where $D_{i,\text{ref}}$ is the equivalent diversity of each thread's first full window. A *contraction episode* begins when three consecutive valid windows have $r_{ij} \leq 0.8$, and an open episode reaches

recovery when three consecutive valid windows have $r_{ij} \geq 0.9$. Onset and recovery are dated to the first window of the confirming run. Recovery concerns the level of equivalent diversity, not restoration of the original pair set or its original distribution. A further contraction after a recovery is a *relapse*. Exact onset, recovery and duration indices are given in Supplementary Methods S3.

The first episode is used for onset, duration and recovery analyses. Recovery and termination of the observed record without recovery are treated as competing endpoints. A contraction is *persistent* when at least 30 comments follow onset, at least 70% of valid post-onset windows remain at or below 0.8, and the final full window is also at or below 0.8; P_i^{low} denotes the share of valid post-onset windows at or below 0.8. These local terms are distinct from the whole-thread classifications above: density-robust tunnel vision does not imply persistence, and a persistent contraction can occur in a thread of any whole-thread classification.

Contraction depth uses the minimum relative diversity over all valid windows in a thread. Onset-aligned trajectories summarise relative diversity among threads still observed at each lag. Supplementary Methods S3 defines these summaries, regained diversity and the cumulative incidence of the competing endpoints.

3.5 Content and conditions analyses

Position dynamics in blocks

To examine how evaluative positions change during contraction and recovery, we divide each thread into non-overlapping blocks of 30 comments, excluding any incomplete final block from this analysis. Let A_{ik} denote the set of active aspect-sentiment pairs in block k of thread i. We measure pair retention between consecutive blocks as

$$\text{Ret}_{ik} = \frac{|A_{i,k-1} \cap A_{ik}|}{|A_{i,k-1}|}. \tag{4}$$

Mentions in the later block are partitioned into three mutually exclusive sources, whose mention shares S_{ik}^{cont}, S_{ik}^{return} and S_{ik}^{first} sum to one: *continued* pairs, present in both adjacent blocks; *returning* pairs, absent from the immediately preceding block but observed earlier in the thread; and *first-observed* pairs, not previously seen in the thread. All three categories refer to retained mentions within the frozen pair space: first-observed does not denote a newly induced aspect, and temporary absence does not imply that participants have abandoned a position. The exact set definitions, the common-pair entropy used to measure rebalancing among surviving positions, and the dominant-pair quantities are given in Supplementary Methods S4.

Blocks are aligned separately on the first onset and first recovery. For any transition quantity X, the event-related paired change is

$$\Delta X_i^{(e)} = X_{i,0}^{(e)} - X_{i,-1}^{(e)}, \qquad e \in \{\text{onset}, \text{recovery}\}, \tag{5}$$

so each thread is compared with its immediately preceding transition. Because first-observed pairs decline with thread position and density-robust onsets occur later, the raw between-group contrast at onset is complemented by direct standardisation to the density-robust group's absolute block-position distribution.

Opening composition

The opening is the first 50 usable comments, with diversity measured by mean ASPE over the five windows lying wholly within it. Subsequent outcomes use only windows beginning at comment 51 or later. We report associations with subsequent ASPE trends and levels by opening-diversity quartile, Spearman correlation and linear models adjusted for log usable comments, schema size, opening mention density and community. Supplementary Methods S5 gives the outcome definitions and model specifications.

3.6 Statistical inference, robustness and replication

Descriptive estimates use 95% percentile intervals from 2,000 community-cluster bootstrap resamples, with equal point weight for each eligible thread. Interquartile ranges describe variation across threads; regression intervals use community-clustered standard errors. Supplementary Methods S6 specifies aggregation and weighting.

Sensitivity analyses examine the effects of varying the contraction episode thresholds, detection start point and reference diversity. To assess robustness to community composition, we also repeat the analyses while excluding each of the five largest communities in turn. These checks evaluate the sensitivity of the findings to methodological choices and community composition, rather than selecting a preferred set of results (Supplementary Methods S6).

A second-period analysis uses April 2018, selected before its results were computed, and contains 1,035 threads sampled from the communities represented in the analysis sample under the same nominal comment band and usable-comment floor. It shares the relevance gate, classifier and analysis code, with schemas induced for new threads. We report raw and community-standardised estimates to assess recurrence of the patterns, without interpreting differences as isolated calendar-year effects. Supplementary Methods S7 documents sampling and the frozen and newly executed pipeline components.

4 Results

The primary results describe the 2,967 analysis-sample threads from April 2019; the second-period analysis is reported separately. Three quantities carry the whole-thread results: the ASPE trend $\rho_{\mathrm{ASPE},i}$ asks whether diversity falls over the thread, the retained equivalent diversity $R_{D,i}$ asks how much of the opening diversity remains at the end, and the relative diversity r_{ij} asks how far a given window lies below the thread's opening reference. Whole-thread prevalence uses the full sample denominator. Recovery

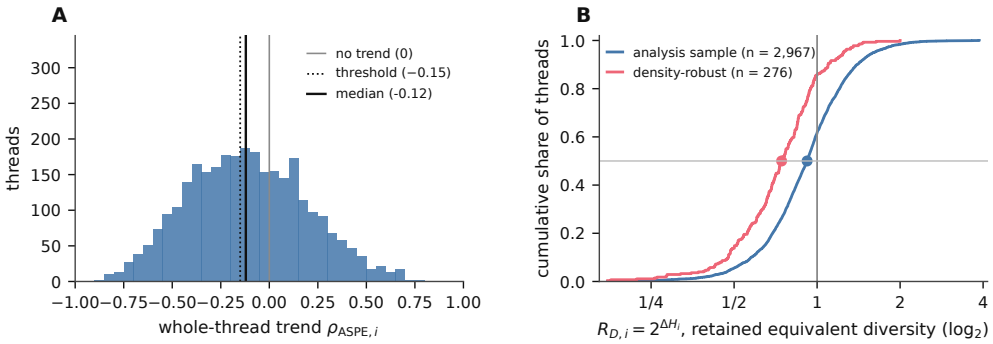


Fig. 2 Whole-thread narrowing. (a) Distribution of the whole-thread ASPE trend $\rho_{\mathrm{ASPE},i}$ (Eq. 2) in the 2,967 analysis-sample threads; dotted line at the −0.15 threshold of observed tunnel vision, solid line at the median. (b) Empirical cumulative distribution of the retained equivalent diversity $R_{D,i}$ (Eq. 3; $\log_2$ axis) for the analysis sample (blue) and the density-robust subset (red); points mark the medians, the vertical line marks no change.

is conditional on an observed first episode, relapse on its recovery, and persistence on an episode with at least 30 comments of follow-up. Block analyses use the valid transitions specified in their captions. Density-robust tunnel vision is compared with all other threads unless stated otherwise.

4.1 Whole-thread narrowing and its robustness

Observed tunnel vision occurs in 45.9% of the analysis sample. Whole-thread ASPE trends are centred slightly below zero (Figure 2a), and declining diversity is closely associated with increasing top-five concentration. The remaining threads do not meet the pre-specified whole-thread criterion, although they may still contain local contractions.

Density-robust tunnel vision occurs in 9.3% of the sample (276 threads). These trends are more negative than expected under the specified null preserving per-comment mention counts and whole-thread pair frequencies. For the other threads with observed tunnel vision, that comparison does not exclude the alternative represented by the null. The density-preserving null itself has a negative mean trend in roughly two thirds of threads, demonstrating why observed narrowing and robustness to mention-density structure are distinct results. The order-shuffle comparison yields 13.7% order-robust

tunnel vision and 6.0% dual-null-robust tunnel vision. The two null-specific subsets overlap without forming a sequence of filters (Supplementary Figure S1 and Table S1).

Narrowing in the density-robust subset is substantial relative to the typical thread. Its median ASPE trend is −0.53, compared with −0.09 among all other threads, and its median retained equivalent diversity is 0.74, compared with 0.92 in the full sample (Figure 2b). Thus its median endpoint retains approximately three quarters of the opening window's equivalent diversity. Individual trajectories vary widely: the density-robust subset's interquartile range of R_D is 0.60–0.92 (Supplementary Table S1). The usable-size bands show no clear monotonic prevalence pattern within the restricted sampling frame (Supplementary Table S2); the supplementary long threads do not support inference about threads beyond the sampling frame.

4.2 Frequent contractions, unequal recovery

Local episodes reveal how diversity changes within the overall trend. We apply the fixed onset, recovery and persistence rules to every thread, irrespective of its whole-thread classification. Density-robust tunnel vision describes robustness of the full trend to a specified null; it does not imply that a local contraction persists to the end.

Contraction episodes occur in 73.8% of threads, and more than half of the threads with an episode experience at least two (Table 1). Temporary losses of diversity are therefore common even outside the density-robust subset. Among threads with an episode, the median minimum relative diversity is similar in the two groups, 0.47 for density-robust tunnel vision and 0.50 for all other threads. The density-robust subset instead has longer runs of low-diversity windows and reaches its minimum later in the observed record.

First onset occurs later in density-robust tunnel vision: at a median 0.39 of thread length, compared with 0.26 among all other threads with episodes (Figure 3a). This is a descriptive feature of the groups, partly related to their classification by the full

Table 1 Contraction episodes by group. Entries are medians with interquartile ranges unless stated; shares carry 95% community-cluster bootstrap intervals; positions and durations are in comments, relative positions are fractions of thread length.

	Analysis sample	Density-robust tunnel vision	All other threads
Threads	2,967	276	2,691
With an episode	2,189 (73.8%; 70.1–76.5)	248 (89.9%; 86.9–93.4)	1,941 (72.1%; 68.2–75.1)
Two or more episodes	1,279	120	1,159
First onset, comments	100 (60–185)	142.5 (80–220)	95 (55–175)
First onset, relative	0.28 (0.16–0.49)	0.39 (0.23–0.63)	0.26 (0.15–0.48)
Minimum relative diversity $r_i^{\min}$	0.50 (0.40–0.58)	0.47 (0.38–0.56)	0.50 (0.41–0.59)
Position of the minimum, relative	0.68 (0.42–0.87)	0.83 (0.66–0.92)	0.65 (0.40–0.85)
Longest low run, comments covered	65 (50–95)	85 (60–125)	65 (50–90)
Longest low run, first to last window	35 (20–65)	55 (30–95)	35 (20–60)
First episode recovered	1,832 (83.7%; 82.0–85.3)	151 (60.9%; 53.9–65.6)	1,681 (86.6%; 85.2–88.1)
Duration to recovery, comments	45 (25–80)	40 (25–75)	45 (25–80)
Relapse after the first recovery	69.8% (68.0–71.9)	79.5% (73.8–85.8)	68.9% (67.0–71.2)
Persisted to the end	323 / 2,146 (15.1%; 13.3–17.2)	85 / 244 (34.8%; 30.4–41.6)	238 / 1,902 (12.5%; 10.9–14.7)
P_i^{low}, share of low windows after onset	0.43 (0.25–0.64)	0.64 (0.44–0.83)	0.40 (0.24–0.60)
End state $r_{i,\text{last}}$	0.84 (0.68–1.02)	0.72 (0.57–0.87)	0.85 (0.69–1.03)
Share of lost diversity regained, G_i	0.67 (0.37–1.03)	0.45 (0.24–0.74)	0.70 (0.40–1.07)

"All other threads" are the analysis-sample threads that are not density-robust. Episode rows refer to threads with an episode; duration to recovery and relapse refer to recovered first episodes; persistence and P_i^{low} use threads with an episode and at least 30 comments of follow-up, and persistence may follow an earlier recovery and renewed contraction.

trajectory. The detection rule also limits early onset: the first window is the reference, so an onset can first be dated at the second window's end and confirmed at the fourth. The earliest bins consequently combine actual early contractions with this detection boundary.

Most first episodes recover within the observed record, but recovery is less frequent in density-robust tunnel vision: 60.9%, compared with 86.6% among all other threads with episodes (Figure 3b; the counts beneath the panels give the first episodes that have reached neither endpoint at each lag). Among those whose first episode recovers,

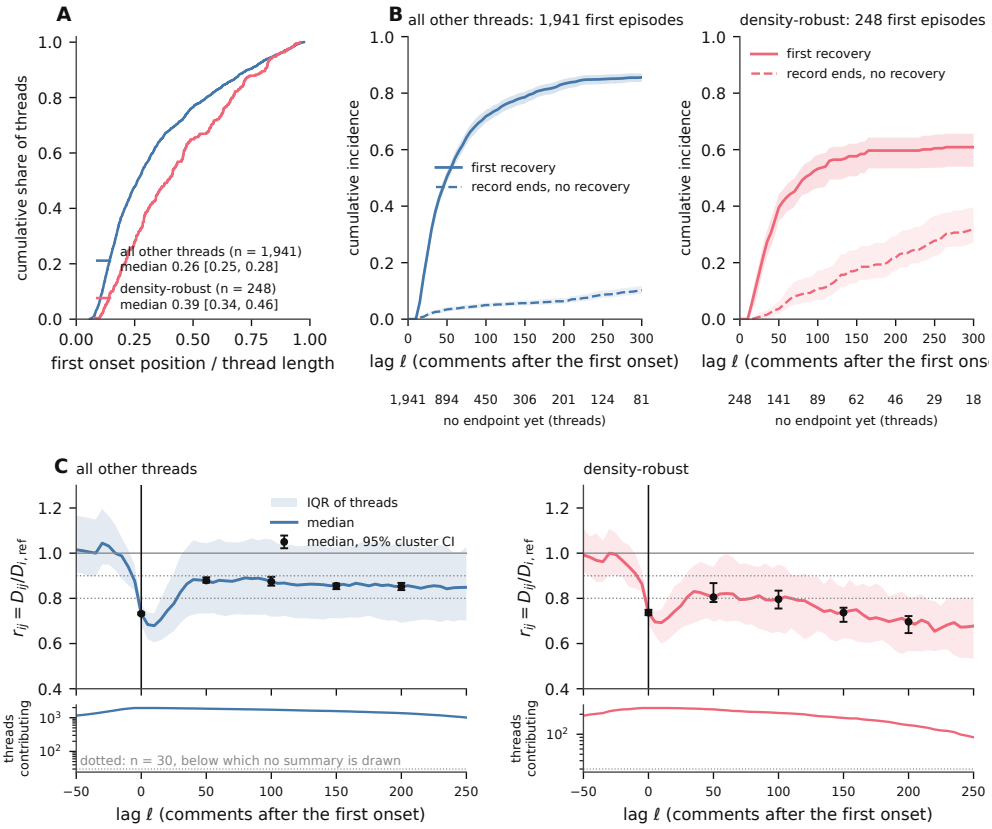


Fig. 3 Onset, recovery and onset-aligned trajectories, for threads with density-robust tunnel vision (red) and all other threads (blue). (a) Empirical cumulative distribution of the relative position of the first onset among threads with an episode. (b) Cumulative incidence, by lag in comments after onset, of a confirmed first recovery (solid) and of the end of the record without recovery (dashed), with pointwise 95% community-cluster bootstrap bands. (c) Relative diversity r_{ij} aligned at the first onset: median line, interquartile band and the median with its 95% cluster interval at fixed lags; lower panels give the number of threads contributing. Dotted lines mark the onset (0.8) and recovery (0.9) thresholds.

renewed contraction is more frequent in the density-robust subset (79.5% versus 68.9%). Persistence to the end is also more frequent (34.8% versus 12.5% among episode-bearing threads with adequate follow-up). Recovery and persistence are not mutually exclusive: a thread can recover and later satisfy the persistence criterion after renewed contraction. The contrast therefore concerns the subsequent course of diversity, not an absence of recovery in every density-robust thread.

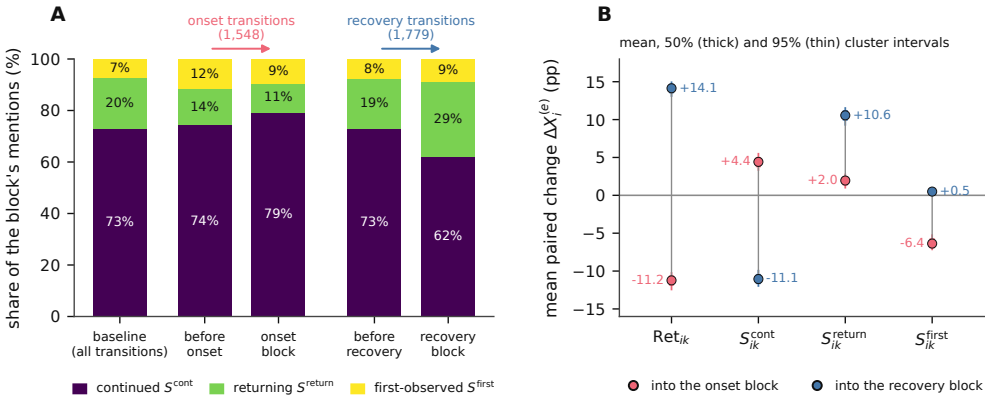


Fig. 4 Mention sources and paired changes around the first episode. (a) Mean shares of a block's mentions from continued, returning and first-observed pairs at the baseline, before and at the onset block, and before and at the recovery block. (b) Mean paired change (Eq. 5) into the onset block (red) and into the recovery block (blue) for the retention of the previous block's pairs (Eq. 4) and for the three mention shares, with 50% (thick) and 95% (thin) community-cluster bootstrap intervals.

Onset-aligned trajectories suggest that the groups differ more in their subsequent recovery than in their initial decline (Figure 3c). Their median trajectories follow similar declines around onset, but all other threads subsequently return closer to the reference level. The density-robust subset shows a weaker rebound and further decline at later lags. These curves describe threads still observed at each lag, with progressively smaller and longer-thread samples toward the right of the plot; the lower panels report the contributing counts, and the summary is drawn only where at least 30 threads contribute. Because both group membership and alignment are derived from the same trajectories, these patterns characterise the form of narrowing rather than establish an independent cause of it.

4.3 Recovery accompanies the return and rebalancing of positions

Pair-level block transitions identify the content changes accompanying contraction and recovery. Each thread is divided into non-overlapping 30-comment blocks, and changes are aligned separately on the first onset (1,548 paired transitions in the analysis sample)

and the first recovery (1,779). Figure 4a shows where a block's mentions come from, continued, returning or first-observed pairs, at the baseline and around the two events; Figure 4b gives the paired changes into the onset and the recovery block; Table 2 adds retention levels, distinct pairs, block entropy and the dominant-pair quantities.

At onset, fewer positions remain represented. Retention of the previous block's pairs falls by 11.2 percentage points relative to the preceding transition, and the mean paired change in distinct pairs is −3.9 (Figure 4b; Table 2). The mentions that remain come more from pairs already present in the previous block, whose share rises from 74% to 79%, and less from first-observed pairs, whose share falls from 12% to 9% (Figure 4a). The current block's most frequent pair takes a larger share, but it need not be the previously dominant pair: the latter loses 1.9 percentage points on average and remains dominant in 45% of onset blocks. Its identity must therefore be distinguished from the maximum share in each block. The median current dominant share is 24%, indicating that the typical onset still involves several positions rather than concentration on a single pair.

Recovery is accompanied by both returning pairs and redistribution among continued pairs. Returning pairs supply 29% of the recovery block's mentions, against 19% at the transition before it and 20% at the baseline (Figure 4a), a paired increase of 10.6 percentage points, while entropy within the common pair set rises by 0.052 bits (Table 2). First-observed pairs contribute a much smaller mean change, +0.5 percentage points, and hold about 9% of the mentions in both event blocks, although they still account for a nonzero share of recovery-block mentions. The paired changes into the two blocks are close to mirror images: retention and the continued share move in opposite directions at onset and at recovery, whereas the first-observed share falls at onset and barely moves at recovery (Figure 4b). The previous dominant pair loses share but remains dominant in 43% of recovery blocks. Renewed diversity thus often

accompanies the return and rebalancing of an existing repertoire; the results do not imply that first-observed pairs are absent or unnecessary.

The two groups share these broad content patterns (Supplementary Figure S3). Their onset-related retention changes are similar, while density-robust tunnel vision shows a weaker rebound in the following block. Returning pairs and common-set rebalancing accompany recovery in both groups. One apparent difference depends on thread position: first-observed pairs become less frequent as threads develop, and density-robust onsets occur later. The raw onset difference in first-observed share is -2.6 percentage points; after standardisation to common block positions it is $+0.1$ percentage points (Supplementary Figure S2). The raw contrast consequently offers little evidence of a separate group-specific difference in the introduction of positions.

The two cases in Figure 5, selected by fixed rules, make these transitions visible with named positions. In the persistent case (Figure 5a), later blocks concentrate on a succession of positions from the existing repertoire as less prominent pairs disappear. In the recovered case (Figure 5b), the leading negative positions remain prominent, but earlier alternatives reappear and the leading positions occupy smaller shares. These cases illustrate the aggregate patterns; their selection does not estimate how frequently either trajectory occurs. All pairs of both threads are shown in Supplementary Figure S4.

4.4 Opening composition and what follows it

Opening diversity is partly embedded in any whole-thread trend. To reduce this overlap, we measure the opening over the first 50 comments and subsequent outcomes only in later windows.

Opening diversity has a small negative association with the post-opening ASPE trend ($\rho = -0.08$), but a substantially larger positive association with late-window ASPE ($\rho = +0.52$). Figure 6a traces the median windowed ASPE after comment 50 by quartile of opening ASPE: all four quartiles decline gently, yet they remain ordered

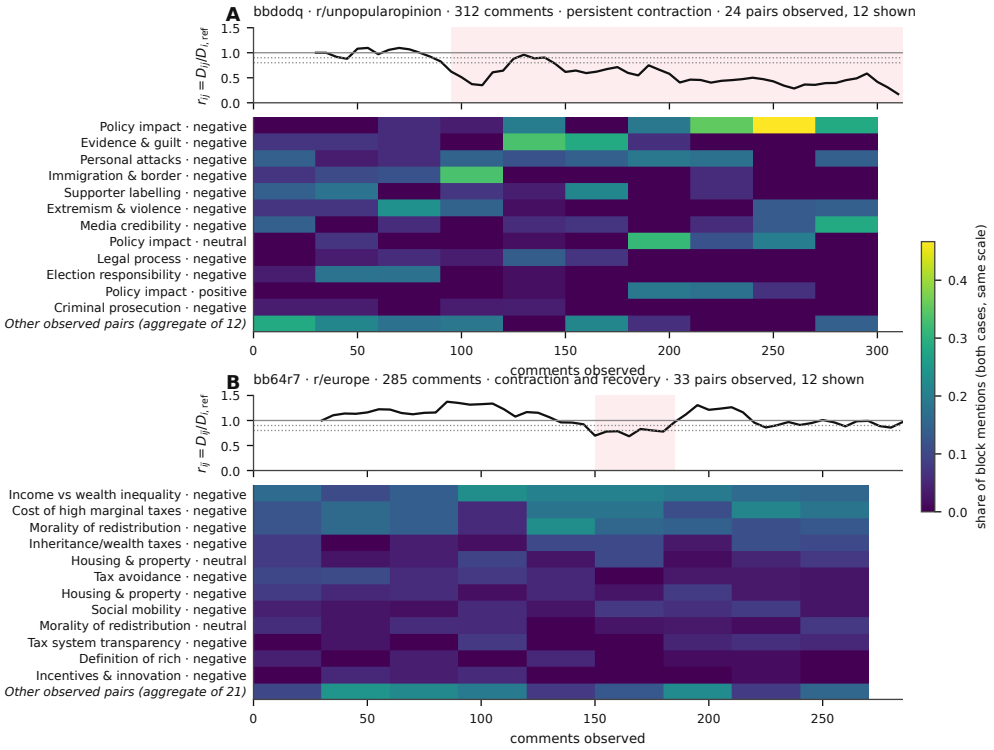


Fig. 5 Two density-robust threads at the level of positions. (a) A thread whose first episode persisted to the end of the record; (b) a thread whose first episode recovered. Top of each panel: relative diversity r_{ij}, with contraction episodes shaded. Bottom: the share of each 30-comment block's mentions held by the twelve most-mentioned aspect-sentiment pairs, with all remaining pairs in one aggregate row.

and approximately parallel throughout the post-opening record, so the gap set in the opening is largely preserved (lines stop where fewer than 100 threads contribute). Figure 6b shows the per-thread distribution of late-window ASPE by opening quartile: the median rises monotonically from 2.99 bits in the lowest quartile to 3.71 bits in the highest, equivalent to 7.9 and 13.1 equally weighted positions, although the distributions overlap substantially. The density-robust share is higher in the top opening-diversity quartile than in the bottom quartile, while late-window diversity also remains higher (Supplementary Table S3). Adjusted models retain the same directions after accounting for thread length, schema size, opening mention density and community (Supplementary Figure S5). Thus a declining trajectory and a low-diversity endpoint describe different

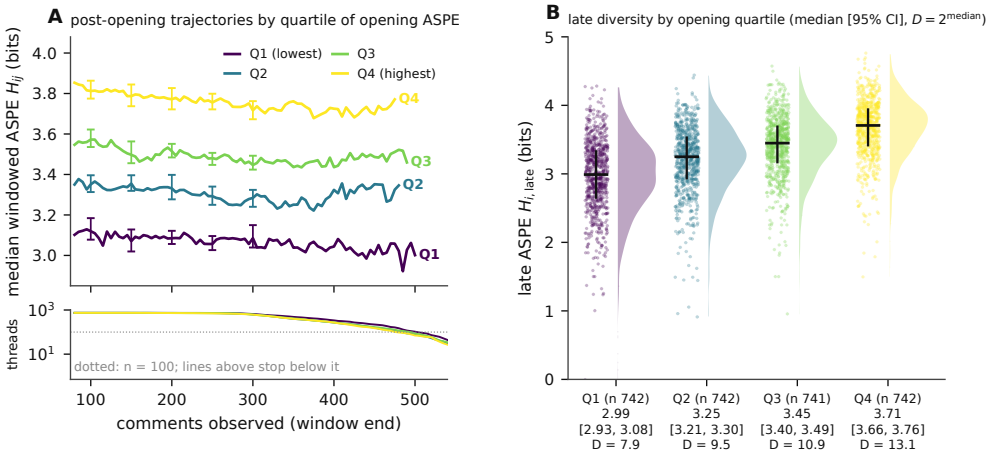


Fig. 6 Opening diversity and later diversity. (a) Median windowed ASPE H_{ij} after comment 50 by quartile of opening ASPE, with the median's 95% community-cluster interval at fixed positions; the lower panel gives the number of threads contributing. (b) Late ASPE $H_{i,\text{late}}$ per thread by opening quartile: points, half-violin, median and interquartile range; tick labels give the median with its interval and the equivalent diversity.

outcomes: discussions that start with more diverse labelled positions often remain comparatively diverse despite subsequent narrowing. Since schemas differ across threads, this comparison concerns entropy within their respective pair spaces, not an identical amount of semantic breadth across events. The ASPE trend is also unrelated to the trend in rarefied lexical diversity across threads ($\rho = +0.002$; Supplementary Figure S6), so evaluative narrowing is not a decline in vocabulary diversity.

4.5 Robustness and a second period

The pre-specified robustness checks concern the descriptive episode rules and community composition; none was used to select the main setting. Episode frequency and timing depend on the onset, recovery and reference rules, as expected for local events, but every alternative definition preserves the direction of the main group contrasts: threads with density-robust tunnel vision narrow later, recover less often and persist more often than all other threads (Supplementary Table S4 and Figure S7). Omitting the five largest communities in turn leaves the key aggregate estimates within the corresponding

full-sample intervals (Supplementary Figure S8). These checks support the qualitative account without treating episode rates as invariant population constants.

The independent April 2018 sample reproduces the central qualitative pattern (Figure 7). Density-robust tunnel vision remains a minority outcome (7.8%, compared with 9.3% in 2019), with the same rounded median retained equivalent diversity of 0.74. Local contractions remain common and most first episodes recover (81.5% versus 83.7%). At recovery, returning share increases by 12.5 percentage points, first-observed share changes by -0.1 percentage points with an interval spanning zero, and common-set entropy rises by 0.036 bits (95% interval $[0.022, 0.055]$). Return and rebalancing therefore recur in both periods.

Standardising to the 2019 community composition over common communities leaves the broad comparison similar: the density-robust estimates are 7.9% and 9.5%. This does not isolate a calendar-year effect. Although the relevance gate, classifier and analysis code are shared, the 2018 schemas were induced afresh and the sampling frame differs in community list and quotas. Schema names and cues affect which mentions pass the gate, and the density-preserving null does not remove these representation differences. Alternative gates on the same 2018 threads are reported in Supplementary Table S5. We interpret the second period as evidence that the temporal and content patterns recur under a documented second-period frame, without attributing the prevalence difference to time alone.

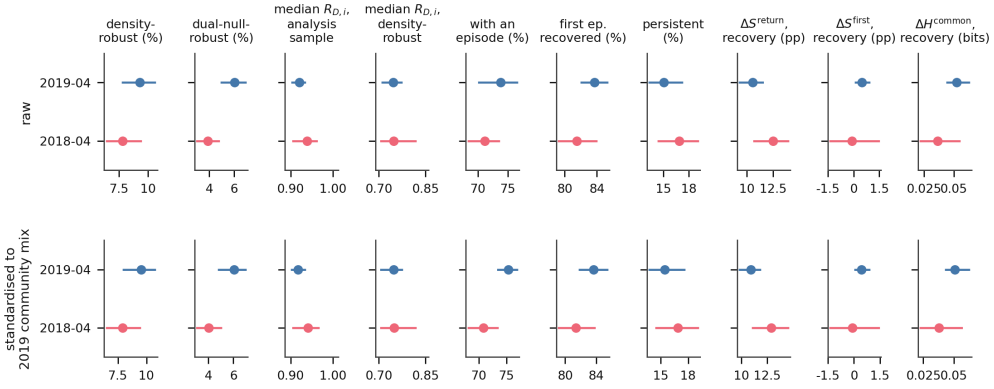


Fig. 7 April 2018 replication. Core estimates in April 2018 (1,035 threads, red) and April 2019 (2,967 threads, blue). Top row: thread-equal-weight estimates with 95% community-cluster bootstrap intervals. Bottom row: the same estimates on the 70 common communities, weighted to the April 2019 community shares.

Table 2 Content transitions around first onset (A) and recovery (B). Levels are medians for retention, distinct pairs, ASPE and dominant share, and means otherwise; paired changes (Eq. 5) are means with 95% community-cluster bootstrap intervals. pp: percentage points.

A. Contraction onset

Quantity	Baseline	Level	Mean paired change (95% interval)
Retention of previous pairs, Ret_{ik}	62.2%	52%	−11.2 pp (−12.4, −10.3)
Distinct pairs	–	10	−3.9 (−4.2, −3.7)
Block ASPE H_{ik} (bits)	–	3.08	−0.51 (−0.53, −0.49)
Dominant-pair share	–	24%	+6.7 pp (+6.2, +7.1)
Continued mentions, S_{ik}^{cont}	72.9%	79%	+4.4 pp (+3.4, +5.5)
Returning mentions, S_{ik}^{return}	20.0%	11%	+2.0 pp (+1.0, +2.7)
First-observed mentions, S_{ik}^{first}	7.1%	9%	−6.4 pp (−7.1, −5.3)
Prev. dominant pair, Δd_{ik} (pp)	–	−1.9	–
Prev. dominant pair still dominant	44.1%	45%	–
$\Delta H_{ik}^{\mathrm{common}}$ (bits)	+0.001	−0.040	–

B. Recovery

Quantity	Baseline	Level	Mean paired change (95% interval)
Retention of previous pairs, Ret_{ik}	62.2%	70%	+14.1 pp (+13.1, +14.9)
Distinct pairs	–	15	+3.9 (+3.7, +4.1)
Block ASPE H_{ik} (bits)	–	3.61	+0.51 (+0.49, +0.53)
Dominant-pair share	–	17%	−7.0 pp (−7.4, −6.5)
Continued mentions, S_{ik}^{cont}	72.9%	62%	−11.1 pp (−12.0, −10.0)
Returning mentions, S_{ik}^{return}	20.0%	29%	+10.6 pp (+9.3, +11.5)
First-observed mentions, S_{ik}^{first}	7.1%	9%	+0.5 pp (+0.1, +0.9)
Prev. dominant pair, Δd_{ik} (pp)	–	−12.3	–
Prev. dominant pair still dominant	44.1%	43%	–
$\Delta H_{ik}^{\mathrm{common}}$ (bits)	+0.001	+0.052	–

Onset contributes 1,548 paired transitions and 2,178 blocks; recovery contributes 1,779 and 1,811. Baselines average valid transitions within thread and then across eligible threads. The final three rows are event-transition means; Δd_{ik} and $\Delta H_{ik}^{\mathrm{common}}$ are adjacent-block differences.

5 Discussion and Conclusions

This study connects whole-thread narrowing to local contractions, recoveries and changes in expressed positions. Temporary contraction is common and recovery frequent, while density-robust tunnel vision remains a minority outcome. Compared with all other threads, those exhibiting density-robust tunnel vision recover less frequently and are more likely to experience relapse following recovery. Recovery is characterised by the return of previously expressed pairs and a more even distribution among continued pairs, while the share of first-observed pairs changes little.

5.1 Conversational diversity as an evolving repertoire

Our findings complement research on echo chambers and collective attention [2, 4] by examining how the range of evaluations expressed within a discussion changes over time. They also add to research on recurring interaction patterns and event-related semantic compression [3, 6]. In particular, participation, lexical diversity and evaluative diversity capture different dimensions of online discourse. A discussion can remain active and linguistically diverse while increasingly concentrating on fewer aspect-sentiment pairs (Supplementary Figure S6).

Representing evaluations as aspect-sentiment pairs is important because neither aspect diversity nor overall sentiment alone captures the full range of evaluative positions. For example, a discussion may continue to address the same aspect while no longer expressing certain sentiments toward it, or retain both positive and negative sentiments while overlooking other aspects. Tracking these pairs over time reveals which evaluative positions remain represented, which disappear from the observed discussion and which return. However, greater evaluative diversity does not necessarily indicate better discussion quality. A discussion may regain diversity through the repetition of inaccurate, misleading or unconstructive evaluations. Our measures capture the

range and distribution of expressed positions, but do not assess their factual accuracy, reasoning quality or contribution to constructive discussion.

Recovery provides further insight into how discussions regain evaluative breadth. Our findings show that recovery is accompanied mainly by the return of previously expressed positions and a more even distribution among continued pairs, while the share of first-observed pairs changes relatively little. These patterns suggest that renewed diversity often draws on positions already expressed within the discussion rather than relying primarily on new ones. They provide a basis for future research into the mechanisms through which discussions maintain or regain evaluative diversity.

5.2 Interpreting narrowing and its practical significance

Distinguishing whole-thread trends from local contraction episodes provides a more complete understanding of how evaluative diversity changes over time. A temporary contraction may be followed by recovery, while a thread exhibiting density-robust tunnel vision may recover before experiencing relapse. Opening diversity adds another important distinction between the direction of change and the final level of diversity. Threads with more diverse openings tend to retain higher final diversity, even when they show slightly stronger subsequent narrowing. Thus, understanding tunnel vision requires considering not only whether diversity declines, but also how much diversity remains and how contraction and recovery unfold throughout a discussion.

These findings also have implications for the design and evaluation of online discussion platforms. Features that resurface previously expressed alternatives or draw attention to overlooked aspects could be evaluated by tracking the return of evaluative positions, recovery in diversity and subsequent changes in evaluative breadth. However, whether such interventions actually broaden discussions or improve their quality remains to be established through further empirical research.

5.3 Scope and limitations

The findings are based on selected Reddit discussions within a restricted platform-reported comment-count range and two monthly archives. The supplementary long threads do not support generalisation to substantially longer discussions. Although the main patterns recur in the second period, differences in sampling and measurement limit direct comparisons of prevalence. Community standardisation accounts for differences in the measured community composition but not for other differences between the two samples.

The findings also depend on the measurement framework. Model-generated aspect schemas and cue-based matching may miss relevant expressions or assign them to inappropriate aspects, while language-model representations of human opinions and behaviour remain imperfect [25]. The human audit shows that identifying whether an aspect is expressed is more challenging than assigning its sentiment. Within annotator-consensus cells, precision is lower in density-robust threads than in all other threads, although estimates are similar between the early and late thirds of density-robust threads (Supplementary Table S6). Disputed cells are excluded from these comparisons, and the audit finds no evidence that recall declines with thread position (Supplementary Results R2). Nevertheless, neither these findings nor the null comparisons rule out measurement errors in the observed diversity trajectories.

Thread-specific schemas introduce an additional limitation. Although schema size has weak associations with whole-thread trends and retained equivalent diversity, entropy levels are more sensitive to schema capacity (Supplementary Results R3). Adjusting for schema size or normalising entropy cannot establish a common semantic scale across discussions, because schema definitions and lexical cues determine which evaluations are recognised and how they are represented. Moreover, because schema induction draws on comments from across each thread, the measurement supports retrospective analysis rather than direct real-time detection.

Finally, the statistical analyses are descriptive rather than causal. The null-based classifications assess robustness to specified randomisation models and do not establish the causes of tunnel vision. Whole-thread classifications and local contraction and recovery measures are derived from the same comment sequences, so group differences describe associations rather than independent effects. Similarly, opening-diversity associations are retrospective and do not establish predictive or causal relationships. Community-cluster bootstrap intervals account for sampling variability within the resampling framework, but not selection bias or systematic measurement error. Episode detection depends on the fixed window, reference diversity and thresholds, while overlapping windows are not independent. Recovery is assessed only within the observed record, and onset-aligned summaries reflect the threads still available at each time point.

Abbreviations. ASPE: aspect-sentiment pair entropy.

Supplementary information. Additional file 1: Supplementary information (PDF). Detailed measurement and statistical methods, data provenance, additional results, and sensitivity analyses (Supplementary Methods S1–S8, Supplementary Results, Tables S1–S6 and Figures S1–S8).

Declarations

Availability of data and materials. This study analysed publicly available Reddit data from the Pushshift dataset (Baumgartner et al., 2020), covering April 2018 and April 2019. Data selection and processing are described in the manuscript, with further details of sample construction, annotation, and validation provided in Supplementary Information.

Competing interests. The authors declare that they have no competing interests.

Funding. This research received no specific funding.

Authors' contributions. XW designed the study and methods, developed the analysis code, prepared the figures and wrote the first draft. YL contributed to the study design, supervised the work and revised the manuscript. JX and YW developed the sentiment pipeline and carried out the human annotation. MN revised the manuscript. WL coordinated the project and revised the manuscript. SW coordinated the project and provided resources. QB contributed to the research idea. All authors read and approved the final manuscript.

Acknowledgements. Not applicable

References

[1] Wang, X., Liu, Y., Wu, S., Zhao, Z., Hu, Y., Li, W., Bai, Q.: Ideological isolation in online social networks: A survey of computational definitions, metrics, and mitigation. Neurocomputing, 134333 (2026)

[2] Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., Starnini, M.: The echo chamber effect on social media. Proceedings of the National Academy of Sciences **118**(9), 2023301118 (2021)

[3] Avalle, M., Di Marco, N., Etta, G., Sangiorgio, E., Alipour, S., Bonetti, A., Alvisi, L., Scala, A., Baronchelli, A., Cinelli, M., *et al.*: Persistent interaction patterns across social media platforms and over time. Nature **628**(8008), 582–589 (2024)

[4] Lorenz-Spreen, P., Mønsted, B.M., Hövel, P., Lehmann, S.: Accelerating dynamics of collective attention. Nature Communications **10**(1), 1759 (2019)

[5] Cui, H., Kertész, J.: Attention dynamics on the chinese social media sina weibo during the covid-19 pandemic. EPJ data science **10**(1), 8 (2021)

[6] Desiderio, A., Mancini, A., Cimini, G., Di Clemente, R.: Highly engaging events

reveal semantic and temporal compression in online community discourse. PNAS nexus **4**(3), 056 (2025)

[7] Wang, X., Liu, Y., Li, W., Bai, Q.: Tunnel vision in online discourse: Formalization and entropy-based quantification with llm-simulated agents. In: Pacific Rim International Conference on Artificial Intelligence, pp. 297–312 (2025). Springer

[8] Pariser, E.: The Filter Bubble: What the Internet Is Hiding from You. penguin UK, London (2011)

[9] Barberá, P., Jost, J.T., Nagler, J., Tucker, J.A., Bonneau, R.: Tweeting from left to right: Is online political communication more than an echo chamber? Psychological Science **26**(10), 1531–1542 (2015)

[10] Bakshy, E., Messing, S., Adamic, L.A.: Exposure to ideologically diverse news and opinion on Facebook. Science **348**(6239), 1130–1132 (2015)

[11] Garimella, K., Morales, G.D.F., Gionis, A., Mathioudakis, M.: Quantifying controversy on social media. ACM Transactions on Social Computing **1**(1), 1–26 (2018)

[12] Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H.E., Quattrociocchi, W.: The spreading of misinformation online. Proceedings of the National Academy of Sciences **113**(3), 554–559 (2016)

[13] Cota, W., Ferreira, S.C., Pastor-Satorras, R., Starnini, M.: Quantifying echo chamber effects in information spreading over political communication networks. EPJ Data Science **8**(1), 35 (2019)

[14] Bail, C.A., Argyle, L.P., Brown, T.W., Bumpus, J.P., Chen, H., Hunzaker, M.F., Lee, J., Mann, M., Merhout, F., Volfovsky, A.: Exposure to opposing views

on social media can increase political polarization. Proceedings of the National Academy of Sciences **115**(37), 9216–9221 (2018)

[15] Medvedev, A.N., Delvenne, J.-C., Lambiotte, R.: Modelling structure and predicting dynamics of discussion threads in online boards. Journal of Complex Networks **7**(1), 67–82 (2019)

[16] Yu, Y., Jiang, J., Dhillon, P.S.: Characterizing the structure of online conversations across reddit. Proceedings of the ACM on Human-Computer Interaction **8**(CSCW2), 1–23 (2024)

[17] Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. Journal of Machine Learning Research **3**, 993–1022 (2003)

[18] Angelov, D., Inkpen, D.: Topic modeling: Contextual token embeddings are all you need. In: Findings of the Association for Computational Linguistics: EMNLP 2024, pp. 13528–13539 (2024)

[19] Zhang, W., Li, X., Deng, Y., Bing, L., Lam, W.: A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. IEEE Transactions on Knowledge and Data Engineering **35**(11), 11019–11038 (2023)

[20] Šmíd, J., Přibáň, P., Kral, P.: Llama-based models for aspect-based sentiment analysis. In: Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, pp. 63–70 (2024)

[21] Bai, Y., Han, Z., Zhao, Y., Gao, H., Zhang, Z., Wang, X., Hu, M.: Is compound aspect-based sentiment analysis addressed by llms? In: Findings of the Association for Computational Linguistics: EMNLP 2024, pp. 7836–7861 (2024)

[22] Gilardi, F., Alizadeh, M., Kubli, M.: ChatGPT outperforms crowd workers for

text-annotation tasks. Proceedings of the National Academy of Sciences **120**(30), 2305016120 (2023)

[23] Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., Yang, D.: Can large language models transform computational social science? Computational Linguistics **50**(1), 237–291 (2024)

[24] Calderon, N., Reichart, R., Dror, R.: The alternative annotator test for llm-as-a-judge: How to statistically justify replacing human annotators with llms. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 16051–16081 (2025)

[25] Hu, T., Baumann, J., Lupo, L., Collier, N., Hovy, D., Röttger, P.: Simbench: Benchmarking the ability of large language models to simulate human behaviors. In: International Conference on Learning Representations, vol. 2026, pp. 28290–28341 (2026)

[26] Brady, W.J., McLoughlin, K., Doan, T.N., Crockett, M.J.: How social learning amplifies moral outrage expression in online social networks. Science advances **7**(33), 5641 (2021)

[27] Shmueli, G.: To explain or to predict? Statistical Science **25**(3), 289–310 (2010) https://doi.org/10.1214/10-STS330

[28] Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., Blackburn, J.: The Pushshift Reddit dataset. In: Proceedings of the International AAAI Conference on Web and Social Media (ICWSM), vol. 14, pp. 830–839 (2020)

[29] He, P., Liu, X., Gao, J., Chen, W.: DeBERTa: Decoding-enhanced BERT with disentangled attention. In: International Conference on Learning Representations (ICLR) (2021)

[30] He, P., Gao, J., Chen, W.: DeBERTaV3: Improving DeBERTa using ELECTRA-style pre-training with gradient-disentangled embedding sharing. In: International Conference on Learning Representations (ICLR) (2023)

[31] Yang, H., Zhang, C., Li, K.: Pyabsa: A modularized framework for reproducible aspect-based sentiment analysis. In: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, pp. 5117–5122 (2023)

[32] Jost, L.: Entropy and diversity. Oikos **113**(2), 363–375 (2006) https://doi.org/10.1111/j.2006.0030-1299.14714.x

[33] Farine, D.R.: A guide to null models for animal social network analysis. Methods in Ecology and Evolution **8**(10), 1309–1320 (2017) https://doi.org/10.1111/2041-210X.12772

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SI.pdf