scientific reports

OPEN

Generation and influence of eccentric ideas on social networks

Sriniwas Pandey[1✉], Yiding Cao[1], Yingjun Dong[2], Minjun Kim[1], Neil G. MacLaren[3], Shelley D. Dionne[1], Francis J. Yammarino[1] & Hiroki Sayama[1,4✉]

Studying extreme ideas in routine choices and discussions is of utmost importance to understand the increasing polarization in society. In this study, we focus on understanding the generation and influence of extreme ideas in routine conversations which we label "eccentric" ideas. The eccentricity of any idea is defined as the deviation of that idea from the norm of the social neighborhood. We collected and analyzed data from two sources of different nature: public social media and online experiments in a controlled environment. We compared the popularity of ideas against their eccentricity to understand individuals' fascination towards eccentricity. We found that more eccentric ideas have a higher probability of getting a greater number of "likes". Additionally, we demonstrate that the social neighborhood of an individual conceals eccentricity changes in one's own opinions and facilitates generation of eccentric ideas at a collective level.

With rapidly diminishing global boundaries, readily available communication devices, and increasing popularity of social media, a new trend of being "extreme" is becoming normal in today's attention-driven society. In domains like religion and politics, this trend is evident, whereas in areas like music, lifestyle, food, healthcare, and other day-to-day choices, it is not as visible but still present covertly. The ubiquitous presence of such trends makes it important to understand their causes and effects and identify methods to intervene when necessary and appropriate.

Religious and violent extremism has been a well-discussed topic in the literature due to their global presence and destructive outcomes[1–5]. However, there are other categories of extremism that grow gradually over time and remain unnoticed until their ultimate consequences are visible. We use the term "eccentricity" to distinguish such extremism from traditional political and religious extremism. The impact of eccentricity can be perceived in various forms. By spreading hate and disharmony, eccentric behavior polarizes society[6,7]. Fanaticism for favorite celebrities, singers, and politicians often leads to personal rivalry, threats, and cyber-bullying[8,9]. In some other scenarios, eccentricity can be more injurious and fatal too. Such outcomes have been observed in anti-vaccine movements[10–12] and firearms-related discourses[13,14]. The impact of eccentric opinions and behaviors is not only limited to social and individual issues, but it influences financial and economic domains too. Industries often utilize this behavioral tendency to change the dynamics of the market. The popularity of pure vegetarian/non-vegetarian food[15,16], vegan diets, and GMO/anti-GMO foods[17] is the result of such exploitation by the related corporations and social organizations. These examples suggest that a systematic investigation and research on this category of extremism (eccentricity) is required.

Understanding the genesis and evolution of extremism has always been of great interest to researchers[18–22]. Factors like gender, race, education, and upbringing do affect the thought process of an individual[23–25]. However, the influence of society, social interactions, and information exposure plays a dominant role in building or altering opinions[21,26–32]. To understand the factors and underlying opinion dynamics researchers have developed different strategies[33,34]. Researchers have applied statistical physics and other mathematical models to examine the dynamics of extreme idea generation[35–37]. Simulation models like agent-based modeling have also been utilized in several studies[21,36,38,39]. In this study, we uniquely investigated this problem using multiple empirical data of human idea generation and evolution: experimental data obtained through human-subject experiments of idea generation[40,41] and online ideation and interaction data obtained from public social media GAB[42]. We analyzed both datasets and compared the results to identify common patterns of how eccentricity may arise in social

¹Binghamton University, Binghamton, NY, USA. ²University of Texas Health Science Center at Houston, Houston, TX, USA. ³State University of New York at Buffalo, Buffalo, NY, USA. ⁴Faculty of Commerce, Waseda University, Tokyo, Japan. ✉email: spandey4@binghamton.edu; sayama@binghamton.edu

networks. To our knowledge, this is the first study that investigated the eccentric idea generation and evolution dynamics using extensive real-world empirical data obtained from diverse sources.

To understand the inception and development of eccentricity, the first step is to establish a quantitative, measurable definition of eccentricity. There are several works that define traditional extremism, but these definitions are restricted to the application and domain of their work and cannot be directly used to define eccentricity in a broader context[43–45]. Each definition covers some aspects of extremism and ignores the rest[46,47]. As most researchers consider extremism in terms of violence, they weigh an opinion based on whether the opinion will result in an action or not[4,45]. If the opinion results in an action (mostly destructive), the opinion is considered extreme, e.g., those leading to terrorism or a coup. The major drawback of this popular definition is that it does not quantify the strength of opinions. As this definition makes a binary classification about extremism or not, it would miss the scenarios where extremism might not result in direct actions but may cause more gradual escalation of opinions.

Another common way to define eccentricity is using the threshold method in continuous models[18,19]. In this method, an opinion's strength is quantified in a range of possible values and if the strength is greater than a predefined threshold, that opinion is considered extreme or eccentric[26,48]. One of the major challenges with this method is deciding the threshold. The threshold value varies for different environments, different backgrounds and different tasks, and the same opinion will not qualify as extreme in one setting versus another. Even in the same setting, with time, the opinions that were initially classified eccentric may sound sensible in presence of more eccentric ideas later, and vice versa.

In our study, we chose the simplest dictionary definition of eccentricity: "deviation from the norm". The norm of a conversation or discussion is defined as the center of all opinions in a social neighborhood, and the eccentricity is quantified as the distance from the norm, both in a semantic metric space that is obtained using machine learning techniques for semantic embedding. This approach makes our definition of eccentricity parameter-free and avoids the problem of binary classification. As the metric is continuous and context-free, the method can be applied to any domain or task without the need for domain-specific knowledge or expertise.

The concept of eccentricity and its quantification, as introduced in this research, holds significant potential for integration into contemporary social media algorithms and various applications where the visibility or impact of individuals or ideas is a critical factor. For instance, in today's social media platforms, content recommendation algorithms rely on metrics like the number of likes or comments to suggest posts and users to others. Similarly, in the context of talk shows or debates, success is often gauged by the volume of applause or views garnered. However, for those seeking to cultivate platforms that prioritize the meaningfulness and relevance of content, the consideration of opinion eccentricity and discourse distinctiveness becomes a valuable tool. By factoring in eccentricity metrics, these platforms can effectively filter out noise and prioritize content that fosters meaningful engagement, ultimately contributing to a more constructive and purposeful digital landscape.

Ideas that become more popular are intrinsically more eccentric

We focus on understanding the influence of the eccentricity of an idea on the amount of attention the idea receives. From two different data sources (human-subject experiments[40,41] and online social media GAB[42]), we collected three types of data: (1) text posts collected from an online experiment for a laptop tagline writing (high collaboration) task, (2) text posts collected from an online experiment for story writing (low collaboration) task, and (3) social media posts from GAB (see "Methods" for details). The number of likes on a post recorded in these datasets is considered a representation of the amount of attention the post received. The eccentricity of a posted opinion is measured by the semantic distance between the idea and the center of all other ideas in the individual's social neighborhood.

The social neighborhood of an individual, u, is his or her ego network, which comprises all other individuals u is following and user u herself. User u gets exposed to the opinions posted in her social neighborhood (Fig. 1a). Recent opinions posted (we used recent 5 days) in the social neighborhood of u constitute a knowledge base of u at each time point t (Fig. 1b). The rationale behind maintaining a knowledge base that contains only recent opinions is rooted in the belief that exposure to contemporary ideas holds a greater potential for inspiring novel insights. Recent opinions tend to align more closely with current trends and prevailing sentiments. Opinions can undergo significant changes, especially within the dynamic realm of social media. For instance, consider the events of April 15, 2019, when the tragic fire at the iconic Notre-Dame Cathedral in Paris became a focal point of discussion on both news outlets and social media. Just a few days later, on April 18, 2019, the release of the Mueller Report on Russian interference in the 2016 U.S. Presidential election took center stage. Within a week, the collective attention on social media shifted from discussions about the Notre-Dame fire to in-depth analysis of the report. Over extended periods, topics and discussions can undergo drastic transformations, often losing relevance and impact on shaping new opinions. By prioritizing recent ideas in building our knowledge base, we can effectively sidestep outdated information and extraneous noise, thereby enhancing the model's efficiency.

In the next step, we convert text ideas in the knowledge base into numerical vectors using the Doc2Vec method[49]. Principal component analysis is performed on these numerical vectors to reduce the dimensionality (see "Methods" for details). The knowledge base is different for each individual and gets updated when there is a new idea in the neighborhood. The eccentricity of a new opinion $O_u(t)$, posted by u at time t, is measured by the distance of $O_u(t)$ from the center of u's knowledge base at that time t (Fig. 1c) (see "Methods" for mathematical definitions).

Posted ideas are partitioned into different popularity levels according to the number of likes they received. We compare the probability distributions of eccentricity for different popularity levels to find the relationship between eccentricity and popularity. We use the kernel density estimation method with a Gaussian filter[50] to
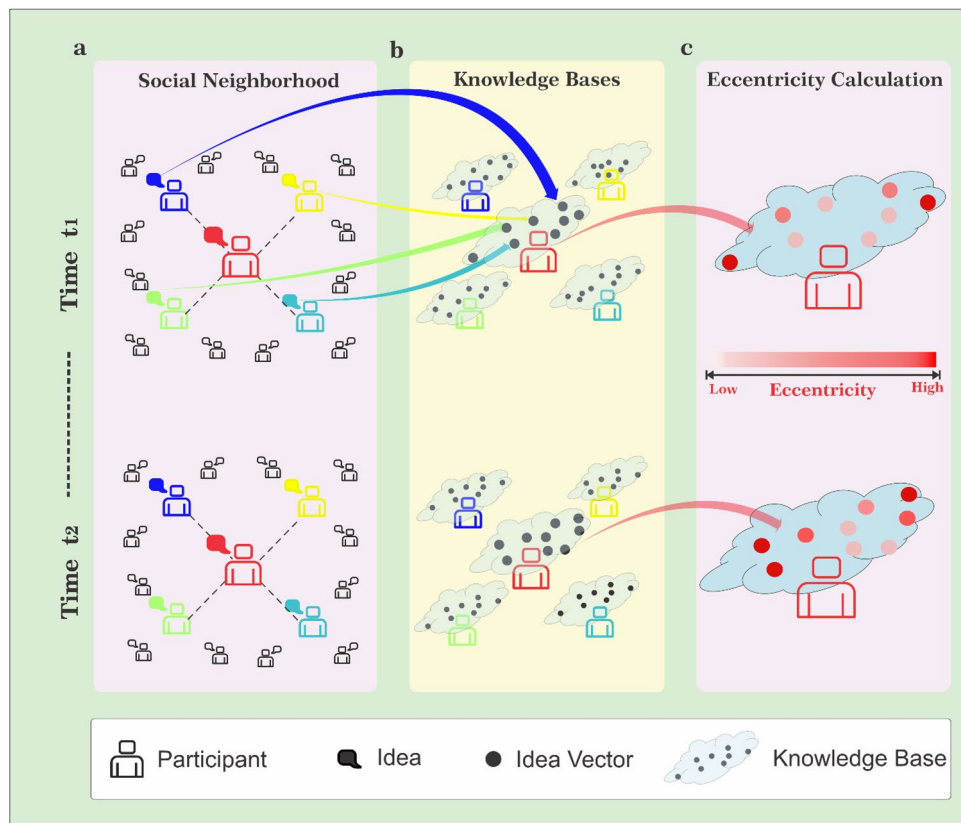
Figure 1. Illustrating the Step-by-Step Process of Eccentricity Calculation Within a Social Network. Participants in a social network are connected to a subset of other participants forming a social neighborhood. (**a**) Each participant can access the ideas generated in the social neighborhood. (**b**) The ideas are vectorized using Doc2Vec. Recent ideas posted in the social neighborhood of a participant (by the participant and their neighbors) constitute the *Knowledge Base* for the participant. (Every participant has a separate knowledge base, which is updated when a new idea is posted in the neighborhood). (**c**) To measure the *eccentricity* of a newly generated idea, distance between new idea (vector) and the center of the knowledge base is measured. The intensity of red color represents the distance from the center of the knowledge base (which is, by definition, the eccentricity).

estimate the probability distribution for each popularity level. Figure 2a and b represent popularity distributions for the posts in the laptop tagline writing task and the posts in the short story writing task, respectively.

The data obtained from the online experiments has a limited range of likes (0–5), hence ideas are partitioned into two popularity levels: **High** (> 2) and **Low** ($<\,= 2$). Figure 2c shows the probability distribution for GAB social media posts. GAB posts, which have a wide range of the number of likes (0 to 500 +), are partitioned into three popularity levels: **Low** ($<\,= 10$), **Medium** (11–100), and **High** (> 100). For all the three datasets, the right tail of the distribution gets broader for the higher popularity levels, indicating that more eccentric posts attract greater attention and popularity. The pattern of increasing average eccentricity with increasing popularity levels is consistent across all the datasets, despite the different nature of the sources. The average eccentricity and eccentricity distributions for different popularity levels are significantly different from each other (p-values obtained by Anderson–Darling test using Bonferroni correction method are shown on each plot).

Movement of neighborhood ideas conceals own ideas' deviation

Furthermore, we propose another variation of eccentricity measure named "self-eccentricity". Self-eccentricity is the eccentricity of opinions with respect to the previous opinions made by the same author of the opinion in question. In other words, the self-eccentricity of an individual's new idea is the distance of the new idea from the center of that individual's previous ideas. Whereas eccentricity is an indicator of deviation from the common consensus or core of the discussion in the social neighborhood at a certain time point, self-eccentricity measures departure from one's own previous ideas over time. We applied this measure to the GAB dataset that had sufficient historical data of users' opinions. We observed that the eccentricities of posts made by an individual did not remain constant but kept changing. We quantified the temporal change of self-eccentricity and analyzed the dynamics of eccentricity change for each user.

In order to quantify the overall evolution of an individual's eccentricity, we have adapted the F-score and G-score metrics proposed by Mall et al.[51] for our analysis. The modified definitions can be found in the "Methods" section. The F-score for a user represents a weighted average of the change in their opinion eccentricity,
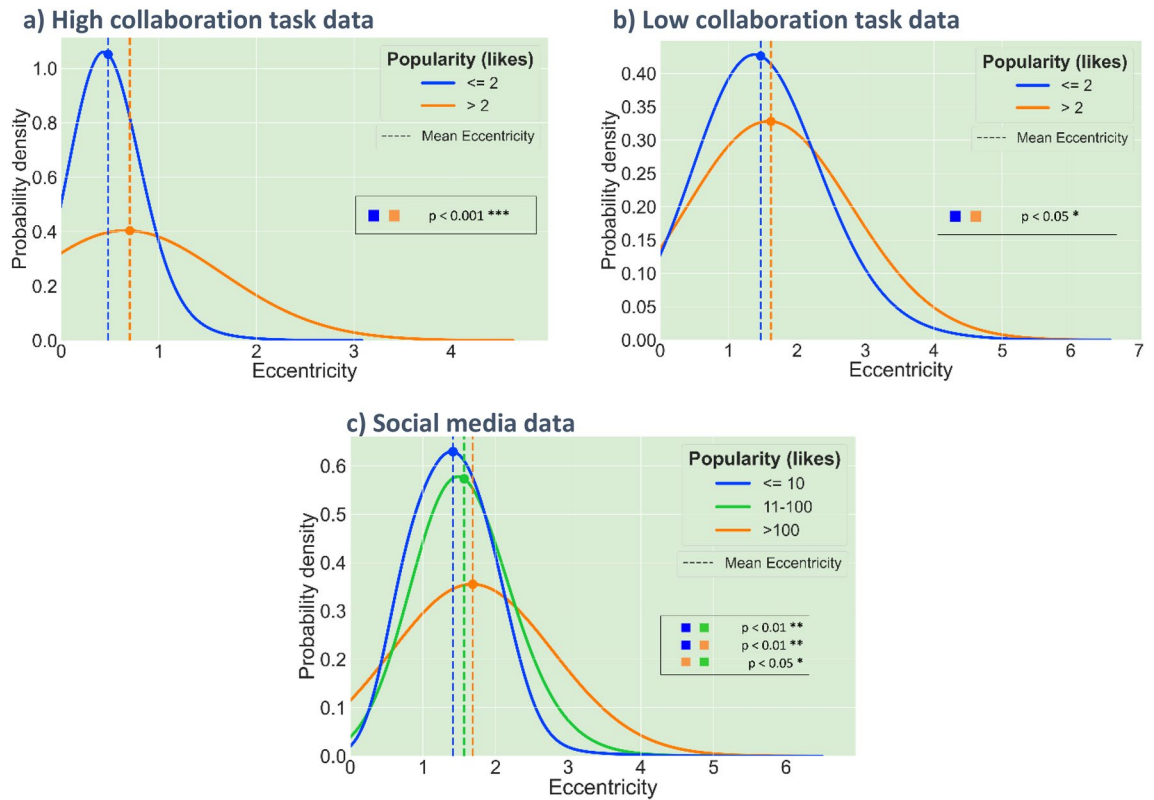
Figure 2. Probability distributions of the eccentricity of the posted ideas at different popularity levels (number of likes, i.e., amount of attention they attracted). Popularity levels are represented in different color bins, low (blue) and high (orange) (and, in case of GAB data, medium (green)). The vertical dashed lines show the average value of eccentricity for each popularity level. (**a**) Plots for the dataset collected in the laptop tagline writing experiment (high collaboration task). Two-sample Anderson–Darling test with Bonferroni correction on unequal sample sizes ($n_{>2} = 59$, $n_{<=2} = 815$) (**b**) Plots for the dataset collected in the short story writing experiment (low collaboration task). Two-sample Anderson–Darling test with Bonferroni correction on unequal sample sizes ($n_{<=2} = 669$, $n_{>2} = 38$) (**c**) Plots for the dataset collected from GAB. Two-sample Anderson–Darling test with Bonferroni correction on unequal sample sizes ($n_{<=10} = 130,234$, $n_{11-100} = 1866$, $n_{>100} 90$). In the plots of high popularity levels, the tail of the probability density function becomes broader in all three data sources. The average eccentricity also increases as the popularity level goes up in all cases.

irrespective of the direction of change. It measures a user's stability in terms of their opinion eccentricity over the period of study. On the other hand, the G-score considers both the change in eccentricity and the direction of that change. It quantifies the extent of average increase or decrease in opinion eccentricity. Both the F-score and G-score definitions incorporate a decay term that factors in the time delay between any two consecutive opinions and the overall average time delay between opinions. This adjustment accounts for changes in an individual's state of mind over time. When combined, the F-score and G-score provide a comprehensive characterization of an individual's behavior. Furthermore, we calculate the F-score and G-score based on the self-eccentricity for each user. The combined F-score and G-score for self-eccentricity help us understand how a user's current opinion deviates from their past opinions.

In Fig. 3a, each individual is represented in a 2-D space based on their F-score and G-score calculated based on the eccentricity (relative to the social neighborhood). The distribution is symmetric, with individuals spread on both the positive and negative sides of the G-score axis. Most individuals are concentrated around the G-score zero line, suggesting that there is no consistent directional trend in idea eccentricity concerning their social neighborhood. In simpler terms, individuals' ideas tend to align with those of their social network.

However, the plot takes on a different pattern in Fig. 3b when users are positioned in the F-score and G-score space of their self-eccentricity. Each user in Fig. 3b is color coded using the same color scheme as in Fig. 3a to show the correspondence between the two plots. Here, the distribution is noticeably skewed toward positive G-scores. This shift indicates that users are generally moving away from their previous ideas, signifying an increase in self-eccentricity over time in relation to their past opinions. This trend holds true regardless of their eccentricity within their social neighborhood, as indicated by the marker colors. The distributions of G-scores for eccentricity and G-scores for self-eccentricity are represented in Fig. 3c. Mann–Whitney test (with an equal sample size) shows a significant difference in these two distributions with G-scores for self-eccentricity higher than G-scores for eccentricity.

These two results, when interpreted together, deliver the following key finding of our work: Individuals are turning more eccentric with time in terms of their own previous ideas (Fig. 3b); however, as everyone in the
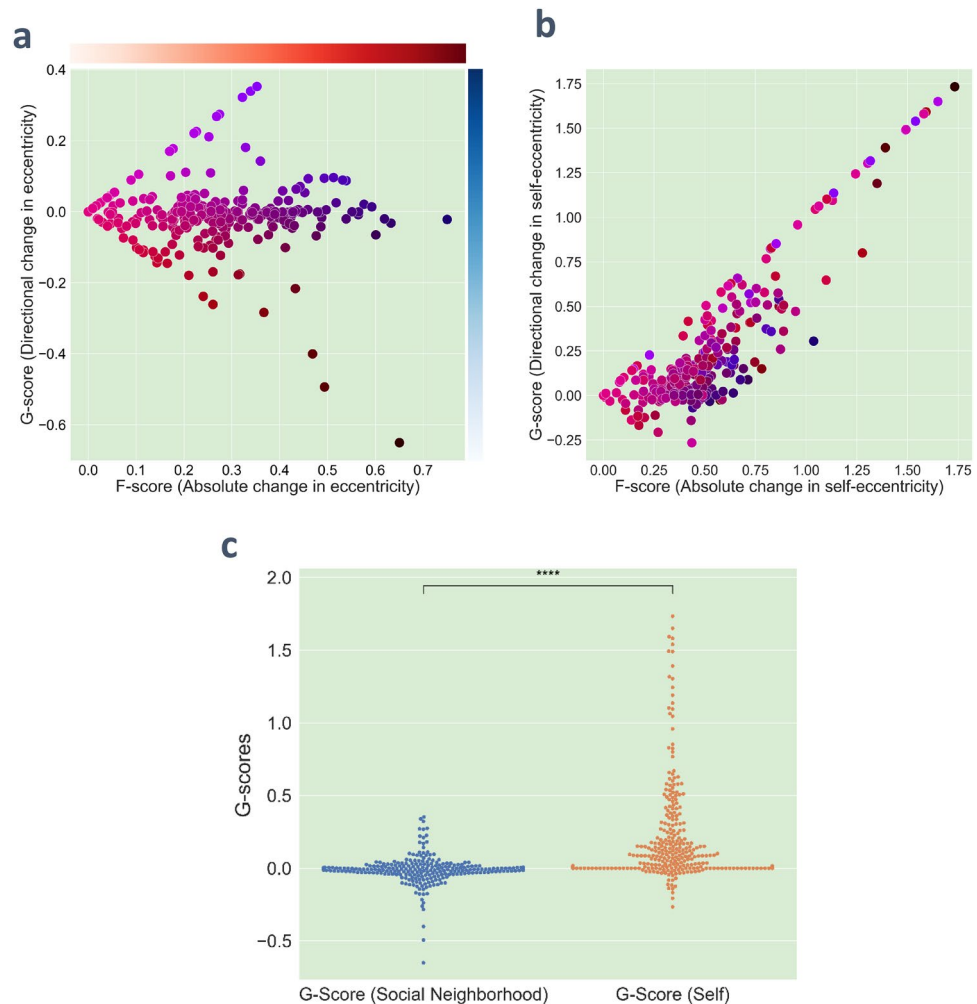
Figure 3. Distribution of GAB users' idea eccentricity dynamics. Dynamics of user opinion eccentricity is explained using two metrics: F-score which quantifies average absolute change and G-score which measures average directional change in idea eccentricities. Each dot represents a user in the data. (**a**) F-scores and G-scores of users' eccentricities in their respective social neighborhood, i.e., deviation of their ideas from the center of their neighbors' ideas. Users are color coded according to their F-score, G-score (calculated based on social neighborhood eccentricity) coordinates as shown in color bars. (**b**) F-scores and G-scores of users' self-eccentricities, i.e., deviation of their idea from their own recent ideas. Each user in (**b**) is color coded using the same color scheme as in (**a**) to show the correspondence between the two plots. (**c**) Distributions for eccentricity G-score (relative to the social neighborhood) and self-eccentricity G-score (relative to own past). The G-score distributions are significantly different (Mann–Whitney test with equal sample size of 318) from each other. The G-score (social neighborhood) mean is lower than the G-score (self-eccentricity) mean (-0.0092 vs 0.201).

neighborhood of these individuals is also shifting from their prior opinions, the change in individual's ideas may not be noticeable (Fig. 3a). Such spontaneous yet unrecognized increase of idea eccentricity can be driven by the positive correlation between eccentricity and attention described earlier (Fig. 2).

Discussion

In today's heavily interconnected world, we notice a trend that a large section of society is increasingly opting for eccentric choices that stand out. We have explored a few insights about this behavior with multiple real-world empirical data. Our first finding shows that the deviation of opinions from the norm helps attract the attention of other individuals. Several studies have shown that serious adverse effects of social media usage are cravings to receive social acceptance and attract the attention of friends and acquaintances[52–54]. Our finding indicates that such human desire may naturally lead to generation of more eccentric opinions. This behavior may scale up to other contexts in the real world beyond online social media.

Another crucial implication about eccentricity obtained in this study is that the overall collective shift of ideas in our social neighborhood may create an illusion of consistency in our own opinions. This can be understood using an analogy of multiple passengers riding on an elevator. In a smooth-moving elevator, any change in elevation is not felt directly by the people using the elevator because the only reference points to assess one's position

are fellow individuals riding on the same elevator. As everyone is moving in the same direction at the same speed, it feels like everyone is standing still and not moving. Similarly, we may not feel the shift in eccentricity of our opinions as our social neighbors show similar shifts.

This study draws a picture of how extreme ideas and opinions may spontaneously arise in society. Everyone wants to gain social acceptance and become popular and influential. As being eccentric in opinions helps attract neighbors' attention, people start expressing out-of-center opinions. And as most of the social neighbors do the same, it would be difficult to notice that one's opinions are becoming more eccentric compared to others. If people do not recognize the heat, there would be little feedback mechanisms to stop them from becoming more eccentric. These behavioral patterns form a cycle and may reinforce each other. These conclusions illustrate the need for further study of how to detect such spontaneous escalation dynamics in society, and if appropriate, how to implement effective interventions so that it will not cause undesirable negative impacts on our lives.

One notably direct and consequential application of this concept is its potential role in the analysis and regulation of our own behavior within the realm of social communication. As demonstrated in our research, the pursuit of attention often leads individuals to express opinions that deviate from the norm. Paradoxically, such behavior, when exhibited by our peers, tends to camouflage our own eccentricities. Without the means to measure and alert us to these behavioral shifts, they may potentially yield adverse long-term consequences.

Drawing from the insights presented in this study, there is an opportunity to develop tools and mechanisms capable of monitoring and notifying individuals about such behavioral changes. These tools could provide individuals with an opportunity to reevaluate and recalibrate their communication patterns. Similarly, such tools can find application within organizational settings to track and analyze communication behavior among members, or in educational contexts to scrutinize student behavior in the classroom, thus affording opportunities for early intervention to prevent situations from becoming irreparable.

Methods

Online experiments and data collection

We collected data from two different sources. One is a human-subject online experiments performed at a mid-sized US university, where students in different majors were recruited to participate in a collaborative textual design task on a Twitter-like online experimental platform[40,41]. Participants were linked to a subset of other participants like a social network setting. There were two types of tasks used for the experiment: (i) writing laptop marketing taglines (high collaboration task) and (ii) writing short fictional stories (low collaboration task). Like in typical social media platforms, participants in these experiments could see only the ideas posted by their social neighbors, like those neighbors' ideas, and add comments to them. Details of the experiments, idea generation process and method of visualizations can be found in these papers[40,41]. All experiments were carried out in accordance with relevant guidelines and regulations and informed consent was obtained from all subjects who participated in the experiments.

The other data source we utilized for data collection is the GAB social media. GAB is a social networking service particularly popular among the far-right people in the US. The data from the GAB is freely available and can be scraped from their website. We used the snowball sampling method to collect data from GAB. We collected around 30 M posts and connections information between authors of these posts. As we are interested in understanding eccentricity, which requires a connected network of individuals, we selected the largest connected component (LCC) from the user network. A subgraph induced by randomly selected 10% of the users in the LCC was used for data analysis to keep the computational demand at a manageable level. This gave us a dataset of around 3,000 GAB users with about 147,000 posts for the analysis. The dataset consists of posts that were made between August 2016 and January 2021.

Text embedding and dimensionality reduction

The first step in analyzing textual data is to convert it into numerical form, a process called text embedding. Before converting text ideas into numerical vectors, each text idea was cleaned to remove stop words, punctuation marks and digits. We also used word stemming to convert different forms of a word to a standard form. In the subsequent step, the Doc2Vec method was used to convert the cleaned text ideas into numerical vectors. Doc2Vec first creates a vocabulary using the text corpus (all ideas combined in this study), trains a model, and infers a numerical vector for each text idea. For the GAB data, we set the inferred vector size to be 300, whereas for online experiment data it was set to 90, given the difference in the data sizes. The 300-dimensional numerical vectors for the GAB data were further transformed to lower dimensional vectors (115 dimensions) using the principal component analysis preserving 90% variance in the data. We call the resulting numerical vector an "idea vector".

Eccentricity and self-eccentricity calculation

To calculate eccentricity of an idea, we first create a social neighborhood for each user, which is a directed network between a user and all other users he or she is following. In our study, we assume the collaboration network remains unchanged for each user during the study. The collaboration network is used to create a knowledge base for each user. A knowledge base is the collection of recent ideas (recent 5 days) made in the neighborhood. The ideas in knowledge base are converted to numerical vectors using Doc2Vec method. The mean vector of a vectorized knowledge base is the center of knowledge base, and the distance of an idea vector from the center is the eccentricity of that idea. It is essential to emphasize that each individual has a unique knowledge base that evolves with the introduction of new ideas in their social neighborhood. Consequently, the eccentricity of a particular opinion is determined based on the individual's current knowledge base at the time when that opinion was generated. This highlights the dynamic and personalized nature of eccentricity assessment.

The mathematical process for determining eccentricity involves a series of steps described as follows: To calculate the eccentricity of an opinion $O_u(t)$ (opinion posted at time t by user u), we need a knowledge base $KB_u(t)$ (Knowledge base of user u at time point t). $KB_u(t)$ is the collection of all opinions in the social neighborhood of user u posted before time t. Social $neighborhood(u)$ contains all the users u is following (or is connected to, in case of undirected network) and user u.

$$KB_u(t) = \{O_v(k) | v \in neighborhood(u), k < t \& (t - k) < 5\, day. \tag{1}$$

Eccentricity of the opinion $O_u(t)$, $Ecc(O_u(t))$ is the distance (d) of $O_u(t)$ from the center of $KB_u(t)$,

$$Ecc(O_u(t)) = d(O_u(t), center(KB_u(t))). \tag{2}$$

Self-eccentricity is the deviation from the center of one's own recent previous ideas (recent 5 days). For each individual, a self-archive of recent ideas is maintained, and each new idea is evaluated against the self-archive. Self-archive $SA_u(t)$ of user u at time t can be defined as:

$$SA_u(t) = O_u(k) | k < t \& (t - k) < 5\, days. \tag{3}$$

Self-eccentricity ($Self\_Ecc(O_u(t))$) of an opinion $O_u(t)$ is the calculated as the distance of $O_u(t)$ from the center of self-archive of the user u at time t.

$$Self_{Ecc(O_u(t))} = d(O_u(t), center(SA_u(t))). \tag{4}$$

In our work, we have used Euclidean distance (L^2 norm) as the distance metric. We have estimated the center of the knowledge bases and the self-archives with the mean of all the contained vectors.

Popularity levels and eccentricity distribution

Ideas are categorized into different popularity levels based on the number of likes they received. The online experiments had a limited number of participants, and each session ran only for two weeks, so the maximum number of likes is less than 6 for this dataset (laptop taglines data < 5, story writing data < 6)[40,41]. Due to the limited number of data points in online experiment dataset, we have opted to categorize the online experiment data into only two groups to ensure the robustness of our claims: ideas having fewer than or equal to two likes (Low popularity) and ideas having more than two likes (High popularity). In the case of the GAB dataset, the range of number of likes is much wider (0 to 500 +). The range of the number of likes differs significantly between the online experiment data and GAB social media data. To make both datasets comparable, we employ a logarithmic scale to normalize the 'number of likes' attribute in the GAB social media data, resulting in a range of 0 to 2.6 on the logarithmic scale. We then define distinct popularity categories for logarithmic likes [0, 1] (original scale 0–10), (1, 2] (11–100), and > 2 (> 100).

For each popularity level, a probability distribution of eccentricity is constructed. The kernel density estimation method is used to estimate the probability distribution. We have used a Gaussian kernel with bandwidth of five to smooth the curve. The mean value of eccentricity is also calculated for each popularity class.

F-score and G-score calculation

For each user u, F-score ($F(u)$) and G-score ($G(u)$) are calculated by taking a weighted average of their eccentricity changes. For F-score calculation, we only consider the magnitude of change of eccentricity not whether it is increasing or decreasing (Eq. (5)). In the G-score calculation, we retain the sign of change of eccentricity to understand the average increase or decrease of eccentricity (Eq. (6)). The additional decay term accounts for delay between two consecutive opinions ($t_{k+1} - t_k$) and overall average delay between opinions (α). N_u is the total number of opinions posted by the user u.

$$F(u) = \sum_{k=1}^{N_u-1} \frac{e^{\left(-\frac{t_{k+1}-t_k}{\alpha}\right)} \left(\left|Ecc\left(O_u\left(t_{k+1}\right)\right) - Ecc(O_u(t_k)\right|\right)}{N_u - 1} \tag{5}$$

$$G(u) = \sum_{k=1}^{N_u-1} \frac{e^{\left(-\frac{t_{k+1}-t_k}{\alpha}\right)} \left(Ecc\left(O_u\left(t_{k+1}\right)\right) - Ecc(O_u(t_k)\right)}{N_u - 1} \tag{6}$$

Implementation details

Python 3.8 was used to implement all the data analysis procedures of the project. To embed text ideas into numerical space, Gensim Doc2Vec library[55] was used. We have used the Distributed Bag of Words (DBOW) model in the Doc2Vec method. The Doc2Vec parameters are different between the online experiment dataset and the GAB dataset as the size of training data is different. For the GAB dataset, the document vector size is 300 and the min_count parameter is set to 10; meanwhile, for the online experiment dataset, the document vector size is 90 and the min_count parameter is set to 7. The NetworkX library[56] was used to create and maintain collaboration networks.

 7

Human research participants

Experiments were conducted after an approval from the Institutional review board at Binghamton University, NY, USA.

Code availability

Code can be made available upon request by contacting the corresponding author (Sriniwas Pandey). Due to the IRB restrictions supporting data is not available.

References

1. Ferrara, E., Wang, W.-Q., Varol, O., Flammini, A. & Galstyan, A. Predicting online extremism, content adopters, and interaction reciprocity. In *Social Informatics* (eds Spiro, E. & Ahn, Y.-Y.) 22–39 (Springer International Publishing, 2016). https://doi.org/10.1007/978-3-319-47874-6_3.
2. Iannaccone, L. R. & Berman, E. Religious extremism: The good, the bad, and the deadly. *Public Choice* **128**, 109–129 (2006).
3. *What Makes a Terrorist*. (2018).
4. Christmann, K. Preventing Religious Radicalisation and Violent Extremism: A Systematic Review of the Research Evidence. *University of Huddersfield* http://www.justice.gov.uk/downloads/publications/research-and-analysis/yjb/preventing-violent-extremism-systematic-review.pdf (2012).
5. Canetti-Nisim, D., Halperin, E., Sharvit, K. & Hobfoll, S. E. A new stress-based model of political extremism: Personal exposure to terrorism, psychological distress, and exclusionist political attitudes. *J. Conflict Resolut.* **53**, 363–389 (2009).
6. Glaeser, E. L. & Sunstein, C. R. Extremism and social learning. *J. Legal Anal.* **1**, 263–324 (2009).
7. Galam, S. Heterogeneous beliefs, segregation, and extremism in the making of public opinions. *Phys. Rev. E* **71**, 046123 (2005).
8. Galán-García, P., de la Puerta, J. G., Gómez, C. L., Santos, I. & Bringas, P. G. Supervised machine learning for the detection of troll profiles in twitter social network: Application to a real case of cyberbullying. *Logic J. IGPL* **24**, 42–53 (2016).
9. Yardi, S. & Boyd, D. Dynamic debates: An analysis of group polarization over time on twitter. *Bull. Sci. Technol. Soc.* **30**, 316–327 (2010).
10. Kata, A. Anti-vaccine activists, Web 2.0, and the postmodern paradigm—An overview of tactics and tropes used online by the anti-vaccination movement. *Vaccine* **30**, 3778–3789 (2012).
11. Broniatowski, D. A. *et al.* Weaponized Health Communication: Twitter bots and Russian trolls amplify the vaccine debate. *Am. J. Public Health* **108**, 1378–1384 (2018).
12. Dubé, E., Vivion, M. & MacDonald, N. E. Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: Influence, impact and implications. *Expert Rev. Vaccines* **14**, 99–117 (2015).
13. Merry, M. Making friends and enemies on social media: The case of gun policy organizations. *Online Inform. Rev.* **40**, 624–642 (2016).
14. Merry, M. K. Narrative Strategies in the Gun Policy Debate: Exploring Proximity and Social Construction. *Policy Studies Journal* **46**, 747–770 (2018).
15. Kuipers, A., Gorton, M. & Schaer, B. Consumer food sciences: Some theories, models and research methods (using Western Balkan countries as a case study). In *Consumer Attitudes to Food Quality Products: Emphasis on Southern Europe* (eds Klopčič, M. *et al.*) 31–54 (Academic Publishers, 2013). https://doi.org/10.3920/978-90-8686-762-2_3.
16. Arora, A. S., Bradford, S., Arora, A. & Gavino, R. Promoting vegetarianism through moralization and knowledge calibration. *J. Promot. Manag.* **23**, 889–912 (2017).
17. Non GMO salt? Water? Food companies exploit GMO free labels, misleading customers, promoting misinformation—Genetic Literacy Project. https://geneticliteracyproject.org/2015/08/24/non-gmo-salt-water-food-companies-exploit-gmo-free-labels-misleading-customers-promoting-misinformation/.
18. Amblard, F. & Deffuant, G. The role of network topology on extremism propagation with the relative agreement opinion dynamics. *Phys. A Stat. Mech. Appl.* **343**, 725–738 (2004).
19. Deffuant, G., Amblard, F., Weisbuch, G. & Faure, T. How can extremism prevail? A study based on the relative agreement interaction model. *Journal of artificial societies and social simulation* **5**, (2002).
20. Franks, D. W., Noble, J., Kaufmann, P. & Stagl, S. Extremism propagation in social networks with hubs. *Adapt. Behav.* **16**, 264–274 (2008).
21. Sayama, H. Extreme ideas emerging from social conformity and homophily: An adaptive social network model. In *The 2020 Conference on Artificial Life* (ed. Sayama, H.) 113–120 (MIT Press, 2020). https://doi.org/10.1162/isal_a_00349.
22. Sayama, H. Going Extreme without Leaders. https://medium.com/@hsayama/going-extreme-without-leaders-e2f1ea0ba520.
23. Arzheimer, K. & Carter, E. Political opportunity structures and right-wing extremist party success. *Eur. J. Polit. Res.* **45**, 419–443 (2006).
24. Arzheimer, K. Contextual factors and the extreme right vote in western Europe, 1980–2002. *Am. J. Polit. Sci.* **53**, 259–275 (2009).
25. Aly, A., Taylor, E. & Karnovsky, S. Moral disengagement and building resilience to violent extremism: An education intervention. *Stud. Confl. Terror.* **37**, 369–385 (2014).
26. Martins, A. C. R. Mobility and social network effects on extremist opinions. *Phys. Rev. E* **78**, 036104 (2008).
27. Fan, K. & Pedrycz, W. Emergence and spread of extremist opinions. *Phys. A Stat. Mech. Appl.* **436**, 87–97 (2015).
28. Holzhauer, S., Krebs, F. & Ernst, A. Considering baseline homophily when generating spatial social networks for agent-based modelling. *Comput. Math. Organ. Theory* **19**, 128–150 (2013).
29. Sayama, H. Enhanced ability of information gathering may intensify disagreement among groups. *Phys. Rev. E* **102**, 012303 (2020).
30. What the Jan. 6 probe found out about social media, but didn't report - The Washington Post. https://www.washingtonpost.com/technology/2023/01/17/jan6-committee-report-social-media/.
31. Strandberg, K., Himmelroos, S. & Grönlund, K. Do discussions in like-minded groups necessarily lead to more extreme opinions? Deliberative democracy and group polarization. *Int. Polit. Sci. Rev.* **40**, 41–57 (2019).
32. Kong, Q., Booth, E., Bailo, F., Johns, A. & Rizoiu, M.-A. Slipping to the extreme: A mixed method to explain how extreme opinions infiltrate online discussions. *Proc. Int. AAAI Conf. Web Soc. Media* **16**, 524–535 (2022).
33. Dong, Y., Zhan, M., Kou, G., Ding, Z. & Liang, H. A survey on the fusion process in opinion dynamics. *Inform. Fusion* **43**, 57–65 (2018).
34. Acemoglu, D. & Ozdaglar, A. Opinion dynamics and learning in social networks. *Dyn. Games Appl.* **1**, 3–49 (2011).
35. Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591–646 (2009).
36. Rainer, H. & Krause, U. Opinion Dynamics and Bounded Confidence: Models, Analysis and Simulation. *Journal of Artificial Societies and Social Simul.* **5**, (2002).

37. Krueger, T., Szwabiński, J. & Weron, T. Conformity, anticonformity and polarization of opinions: Insights from a mathematical model of opinion dynamics. *Entropy* **19**, 371 (2017).
38. Stauffer, D. & Meyer-Ortmanns, H. SIMULATION OF CONSENSUS MODEL OF DEFFUANT et al. ON A BARABÁSI–ALBERT NETWORK. *Int. J. Mod. Phys. C* **15**, 241–246 (2004).
39. Banisch, S., Lima, R. & Araújo, T. Agent based models and opinion dynamics as Markov chains. *Soc. Netw.* **34**, 549–561 (2012).
40. Cao, Y. *et al.* Capturing the production of innovative ideas: An online social network experiment and "idea geography" visualization. In *Proceeedings of the 2019 International Conference of The Computational Social Science Society of the Americas* (eds Yang, Z. & von Briesen, E.) 341–354 (Springer International Publishing, 2021).
41. Cao, Y. *et al.* Visualizing collective idea generation and innovation processes in social networks. *IEEE Trans. Computat. Soc. Syst.* https://doi.org/10.1109/TCSS.2022.3184628 (2022).
42. Zannettou, S. *et al.* What is Gab: A Bastion of Free Speech or an Alt-Right Echo Chamber. in *Companion Proceedings of the The Web Conference 2018* 1007–1014 (International World Wide Web Conferences Steering Committee, 2018). doi:https://doi.org/10.1145/3184558.3191531.
43. Williford, A. C. Blurred lines: What is extremism. *U. Mich. J. L. Reform* **52**, 937 (2018).
44. Coleman, P. T. & Bartoli, A. Addressing Extremism. (2015).
45. Borum, R. Radicalization into violent extremism I: A review of social science theories. *J. Strategic Secur.* **4**, 7–36 (2011).
46. Sotlar, A. Some Problems with a Definition and Perception of Extremism within a Society. *Policing in central and Eastern Europe: Dilemmas of contemporary criminal justice* 703–707 (2004).
47. Martins, A. C. R. Extremism definitions in opinion dynamics models. Preprint at https://arXiv.org/quant-ph/2004.14548 (2020).
48. Martins, A. C. R. Continuous opinions and discrete actions in opinion dynamics problems. *Int. J. Mod. Phys. C* https://doi.org/10.1142/S0129183108012339 (2011).
49. Le, Q. & Mikolov, T. Distributed Representations of Sentences and Documents. in *Proceedings of the 31st International Conference on Machine Learning* 1188–1196 (PMLR, 2014).
50. Parzen, E. On estimation of a probability density function and mode. *Ann. Math. Stat.* **33**, 1065–1076 (1962).
51. Mall, R. *et al.* Four Types of Toxic People: Characterizing Online Users’ Toxicity over Time. in *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* 1–11 (Association for Computing Machinery, 2020). doi:https://doi.org/10.1145/3419249.3420142.
52. Hernández-Serrano, M. J., Jones, B., Renés-Arellano, P. & Campos Ortuño, R. A. Analysis of digital self-presentation practices and profiles of Spanish adolescents on instagram and TikTok. *N. Appr. Ed. Res.* **11**, 49–100 (2022).
53. McElhaney, K. B., Antonishak, J. & Allen, J. P. "They Like Me, They Like Me Not": Popularity and adolescents' perceptions of acceptance predicting social functioning over time. *Child Dev.* **79**, 720–731 (2008).
54. Choi, M., Panek, E. T., Nardis, Y. & Toma, C. L. When social media isn't social: Friends' responsiveness to narcissists on Facebook. *Personal. Indiv. Differ.* **77**, 209–214 (2015).
55. Řehůřek, R. & Sojka, P. Software Framework for Topic Modelling with Large Corpora. in *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* 45–50 (ELRA, 2010).
56. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring Network Structure, Dynamics, and Function using NetworkX. In: Varoquaux, G., Vaught, T. & Millman, J. (eds) *Proceedings of the 7th Python in Science Conference.* 11–15 (2008).

Acknowledgements

Author contributions

S.P. and H.S. conceptualized the study and wrote the manuscript. S.P. carried out implementation of study, social media data collection, data analysis, data visualization and figures preparation. Y.C., Y.D., M.K., N.G.M, S.D.D., F.J.Y. and H.S. contributed to data collection from human experiments. All authors contributed to manuscript review.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.P. or H.S.

Reprints and permissions information is available at www.nature.com/reprints.

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