



OPEN Terrorist attacks sharpen the binary perception of “Us” vs. “Them”

Milan Jović¹, Lovro Šubelj², Tea Golob^{3,4}, Matej Makarovič^{3,4}, Taha Yasseri^{5,6},
Danijela Boberić Krstičev¹, Srdjan Škrbić^{1,4} & Zoran Levnajic⁴✉

Terrorist attacks not only harm citizens but also shift their attention, which has long-lasting impacts on public opinion and government policies. Yet measuring the changes in public attention beyond media coverage has been methodologically challenging. Here we approach this problem by starting from Wikipedia’s *répertoire* of 5.8 million articles and a sample of 15 recent terrorist attacks. We deploy a complex exclusion procedure to identify topics and themes that consistently received a significant increase in attention due to these incidents. Examining their contents reveals a clear picture: terrorist attacks foster establishing a sharp boundary between “Us” (the target society) and “Them” (the terrorist as the enemy). In the midst of this, one seeks to construct identities of both sides. This triggers curiosity to learn more about “Them” and soul-searching for a clearer understanding of “Us”. This systematic analysis of public reactions to disruptive events could help mitigate their societal consequences.

On the evening of 13 November 2015, three suicide bombers struck outside Stade de France in Paris. This was the beginning of the deadliest event in France since the Second World War, culminating in a mass shooting at the Bataclan theatre later that evening. The terrorist attacks claimed the lives of 137 people including the seven attackers. Events such as these prompt debates infused with anger, accusations, and hatred, but also sometimes followed by self-reflection, introspection, and soul-searching^{1–3}. Amid such turmoil, the public focus suddenly shifts to topics and issues that were marginal before the event and are not even necessarily related to it directly. Events in Paris that night sparked curiosity about Islam and its history, Western colonial legacy, and tourist spots in the Middle East, but also triggered discussions about West and East, Christianity and Islam, and “Us” and “Them”^{2, 4–6}. This confrontation has a regrettably violent history, which includes notable incidents such as the Twin Towers attack in 2001 and the London Underground attack in 2005.

Here we ask: Which topics do come to the forefront of the public discourse in the wake of a terrorist attack? Are there any common patterns across various attacks? How long does it take for public attention to return to routine concerns? Past research in this direction has mostly focused on the psychology and politics behind the terrorist attacks and their portrayal in the media^{7–17}, assuming that media coverage reflects the overall public reaction. However, the topics discussed in the media are not necessarily a complete *répertoire* of issues that people will become interested in^{18, 19}. Other quantitative work utilized election results as a proxy measurement, routinely finding a shift to the political right^{20–24}. While terrorism surely affects the voting decisions, it is hard to separate its real effects from the background political context. Moreover, most such studies are naturally limited to specific countries and specific times. Effects of terrorism on economics were examined, often focusing on macro level^{25–28}. Medium and long term shaping of public opinion after terrorist events was also studied^{29, 30}, but investigation of immediate and direct effects is still missing.

To fill this gap, one would seek to identify what topics are searched online more frequently after a terrorist attack. However, the diversity of the search terms concerning any subject makes it difficult to use the search engine statistics for gauging the public interest without having a predefined list (*répertoire*) of keywords. Instead, we employ the viewing statistics of the Wikipedia content. Researchers already shown the correlation between Wikipedia view statistics and Google search volume^{31, 32}, making it a useful proxy measure of the Web public attention³³. Wikipedia viewing statistics has in fact become an established instrument, enabling not only to comprehend but also to predict a broad variety of social phenomena^{19, 34–47}. It can safely be argued that Wikipedia encompasses “all” issues of potential interest, including those ignited by terrorist attacks. As its information

¹Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad, Novi Sad, Serbia. ²Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia. ³School of Advanced Social Studies, Nova Gorica, Slovenia. ⁴Faculty of Information Studies in Novo Mesto, Novo Mesto, Slovenia. ⁵School of Sociology, University College Dublin, Dublin, Ireland. ⁶Geary Institute for Public Policy, University College Dublin, Dublin, Ireland. ✉email: zoran.levnajic@fis.unm.si

is organized into articles, it is convenient for mapping the public responses into certain topics and categories. Articles represent the missing repertoire.

We studied the viewing statistics of all articles in English Wikipedia around the time of 15 terrorist attacks in Europe (see “The sample of terrorist attacks” and Appendix A). By detecting the spikes in *daily viewcount* of specific articles we identified the issues that the public interest is directed towards as a consequence of these events (see “Data preparation”). Unfortunately, this data is available only from 2015 on. This makes the Paris attack of 2015 the first event to commence our analysis.

Results

Changes of collective attention around Paris 2015 attack. We commence by a thorough examination of the Paris attack in 2015, which is the initial event within our sample. We will broaden our analysis to encompass all attacks later. Focusing on the 12 days surrounding the Paris event, we identify 17 articles that received the most attention in this period, shown in Fig. 1 (for details on the identification method see “Sampling the articles around the time of Paris 2015 attack”). Among these 17 articles, we found that five were directly related to the attack while the remaining 12 were not. On the 12 November 2015, a day before the attack, the largest attention was devoted to three articles related to entertainment: Prem Ratan Dhan Payo (Hindi romantic movie released on 12 November), Spectre (2015 James Bond movie released on 26 October), and Fallout 4 (computer game released on 10 November). The attack starts on the evening of 13 November in Paris local time (Since this was still early afternoon in North America, many English-speaking views registered on 13 November, even though the event lasted until the early hours into 14 November.).

This immediately diverts attention to the newly created article *November 2015 Paris attacks* in addition to *The Islamic State of Iraq and the Levant* (ISIL, the main perpetrator) and *Eagles of Death Metal* (rock band that was playing in Bataclan theatre). The attention to these three articles spiked on 14 November, but without dramatically reducing the attention on the above mentioned entertainment articles. On 15 November the collective attention splits among these three attack-related articles, the final fight of the women’s Ultimate Fighting Championship (held on 14 November), and the British TV personality Lady Colin Campbell due to her appearance in a UK reality show on the same day.

The situation is largely unchanged on 16 November, except that attention peaks for other two attack-related articles: *Abu Bakr al-Baghdadi* (leader of ISIL) and the group *Anonymous*, who following the attack announced a major hacking operation against ISIL. Meanwhile, the attention to three main attack-related articles declined and is overshadowed on 17 November by the news that the actor Charlie Sheen was HIV-positive. Eventually, on 20 November, attention is taken by three entertainment releases: the album “25” by the popular singer Adele, the detective TV series Jessica Jones, and the new part of the dystopian movie Hunger games.

Of the five attack-related articles, by 23 November most attention is devoted to ISIL. The viewing dynamics of all articles exhibit an intense spike immediately after the triggering event followed by a gradual decay in days to follow, consistently with^{43,44}. This pattern seems not to differ between attack-related and other articles.

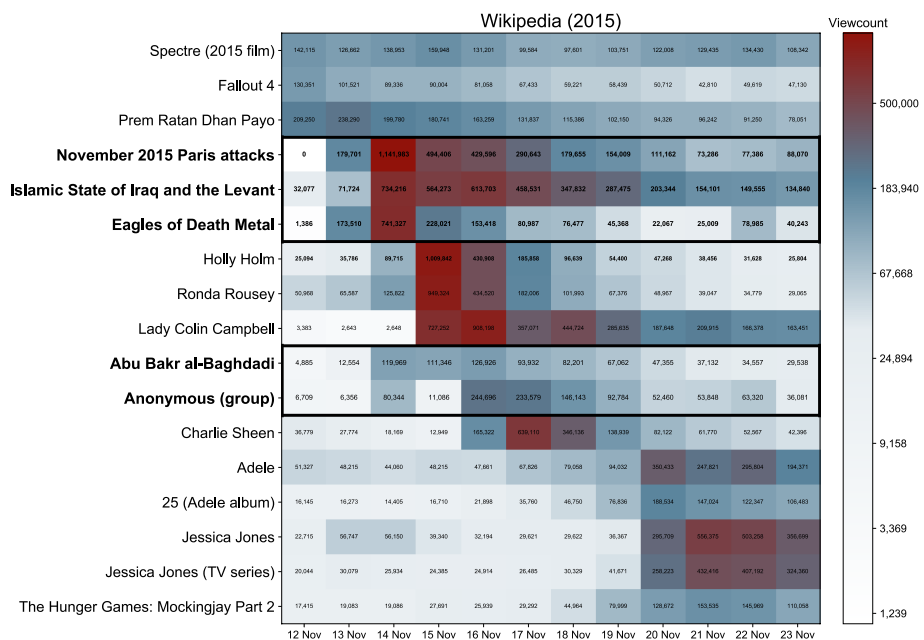


Figure 1. Articles identified with excess attention around the date of the Paris attack in 2015. The color encodes the total views of a given article on a given day (also written within each cell). Articles directly related to the attack are shown in boldface font.

Collective memory of terrorist attacks. We now proceed to assess all 15 attacks in our sample, rather than exclusively focusing on Paris 2015. Research has shown that current events not only attract collective attention to themselves but also trigger collective memories of similar past events^{42–47}. To examine this for the case of terrorist attacks, we look at the Wikipedia articles devoted to all attacks in our sample and ask how strongly each new attack triggers the attention to the previous attacks. In the top left panel of Fig. 2 we show the viewcount time series for Magnanville 2016 from March 2017 through June 2017. Five terrorist attacks that happened during that interval are indicated. Manchester 2017 made the public intensely recall the events of Magnanville 2016. London Bridge 2017 and Westminster 2017 did so somewhat less intensely, while St. Petersburg 2017 and Stockholm 2017 did not at all. Considering this along with the results from the other two left panels, we note that when a terrorist attack happens, the public is reminded of prior (similar) attacks, albeit of some more and some less.

To study this collective memory more systematically, we focus on a quantity we call “excess attention” denoted with Z . It expresses how strongly a newer attack triggers the attention to an older attack (see “Definition of excess attention”). The calculated values of the excess attention are shown in the right panel of Fig. 2. Most events have triggered considerable excess of attention to the past events, confirming the existence of a collective memory of terrorist attacks in Europe. The strongest reactions are found for pairs of attacks that happened in the same country in line with the findings of García-Gavilanes et al.⁴². Also, the time elapsed since the attack seems to be a strong factor, as attacks that happened a long time ago are seldom remembered.

General patterns in public reaction across attacks. Reaction to a terrorist attack is shaped by many factors, primarily by the media portrayal of the event, but also by the perceptions established in the target society about itself and the perpetrators. Searching Wikipedia is a matter of individual choice, but this choice is founded on an array of social narratives that entail symbols, stories and memories of a given group, and turns them into a shared conceptual system^{48,49}. What narratives underpin the searching choices in Wikipedia after a terrorist act? To answer this, we identify the universal effects that attacks have on the collective attention.

Starting from all Wikipedia articles (“Data preparation”), we deployed an exclusion pipeline combining time series analysis and qualitative approaches (“Definition of excess attention” and “Article exclusion steps”), to extract 69 articles to which attention is consistently drawn after terrorist attacks. We then manually grouped them into 19 specific categories according to the commonalities of their thematic codes (“Construction of specific categories”). Finally, by finding coherent patterns among specific categories we grouped them into four broad categories (“Construction of broad categories”). This arrangement is illustrated in Fig. 3 and schematized in the Appendix B. It can be considered in a hierarchy, whereby individual articles belong to one of the specific categories, and each specific category belongs to one broad category. Broad categories represent a coarse, yet robust and interpretable level of description. They are the four core overarching groups of interests that always arise after terrorist attacks.

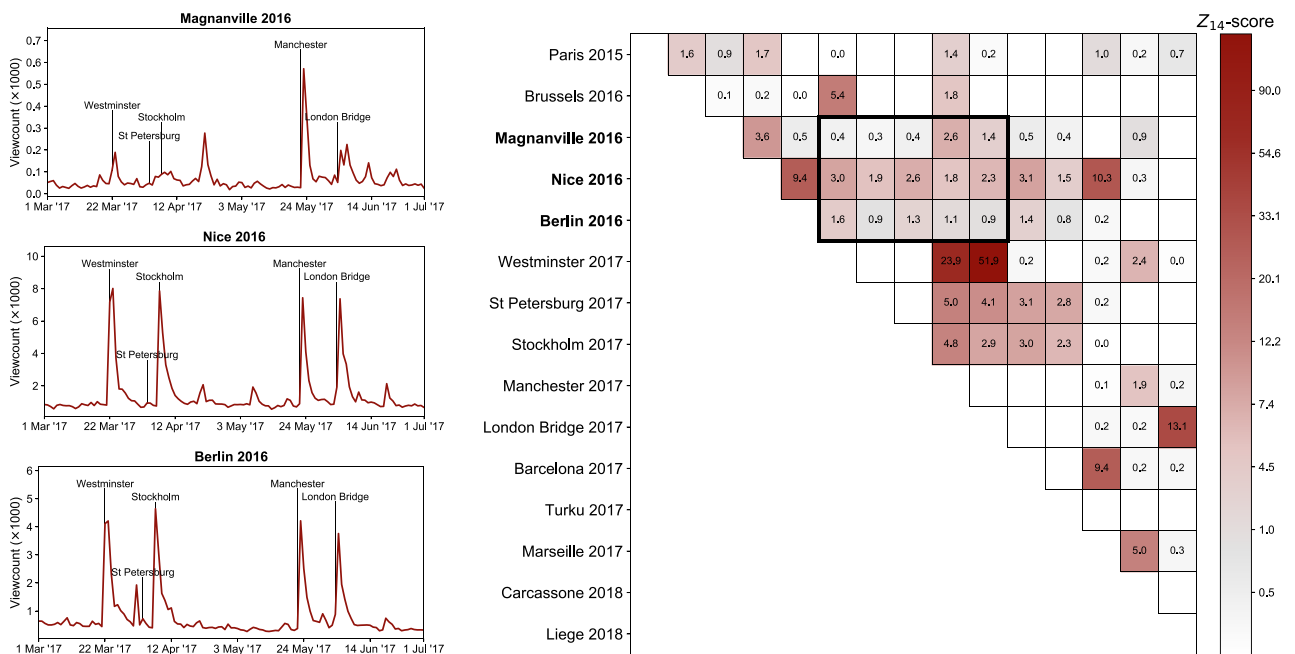


Figure 2. Collective memory of terrorist attacks. Three left panels: time series of viewcounts to articles for Magnanville 2016 (top), Nice 2016 (middle), and Berlin 2016 (bottom), shown for the period of 4 months from March to June 2017. The dates of five terrorist attacks that happened during that time interval are indicated. Right panel: the excess attention Z was received by each attack as triggered by later attacks. Attacks are arranged vertically (and horizontally) in chronological order from top to bottom (left to right). Values of Z for the three panels on the left panel are marked by a bold square.

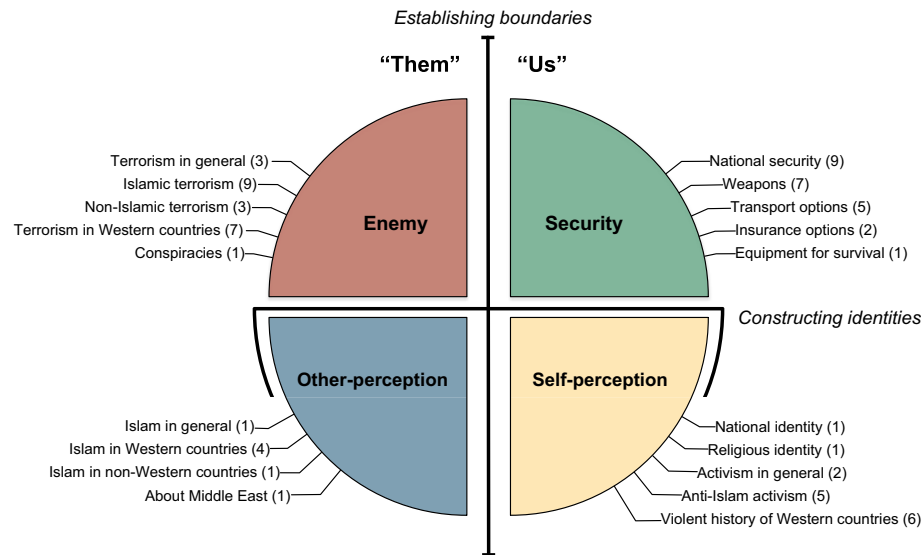


Figure 3. Organization of 69 articles that survived all the filters (Table 1) into specific and broad categories (Appendix B). Four broad categories are illustrated by different colors and positioned as to capture the relationships among them, i.e., the distinctions along two independent dimensions: Establishing boundaries and Constructing identities. Two broad categories on the left explain “Them”, while the two on the right explain “Us”. Specific categories are shown schematically by their names, whereby in parentheses we report the number of articles belonging to each specific category.

Enemy	It is easy to recognize that the concept of <i>enemy</i> , identified as a threat, underpins specific categories related to both terrorism and conspiracies. The basic construction of the enemy goes in two ways: the Islamic extremist, a person driven by extremist ideology posing a danger and/or the terrorist, an individual who is relying on terror as the method of achieving goals in the West and elsewhere. We interpret this as a natural reaction of the English-speaking public to a terrorist attack: Establish the perpetrators (Islamic extremists) as the enemy, and then learn about their methods and actions, especially in the West.
Security	National security, weapons, transport, insurance, and survival equipment have the idea of security as the common denominator. Seeking shelter is a natural response to the threat of the enemy. Being under attack puts us into <i>fight or flight</i> mode: either to confront the enemy or escape from the threat. Fight mode corresponds to the interest in armed forces, military equipment, police, conscription, and usage/purchasing of weapons. The public wants to rely on the official national security mechanisms, but also be ready for individual defence (“take matters into their own hands”). Flight mode is found in the interest of means of transportation (ways to escape), insurance (financial and otherwise), and equipment for survival in the wild (the ultimate shelter in disruptive circumstances).
Other-perception	Once established who is the enemy and how to defend/escape from it, one wants to learn <i>about</i> the enemy. We recognize this as the common motif in pronounced interest for Islam and Muslims, which West often perceives ⁴ as the archetype of the “Other”. Of course, this role has deep historical and cultural roots, which quickly spring out upon a terrorist attack. Induced by it is also a curiosity about the practices of Islam, its origins and history, sacred texts, religious figures, presence in the West and elsewhere, and even tourist attractions in the Middle East.
Self-perception	Finally, one wants to soul-search and re-evaluate who “We” are. This establishes a self-perception that contrasts the perception of “The Other” and relies on both historical and imagined features of the West. It is centered around Western national/religious identities and political activism but also includes anti-Islam groups and figures. Interestingly, there is a large curiosity about violent historic events orchestrated by the Western powers. Given the context of the terrorist attack, it is not surprising that self-perception is built upon (equally) violent aspects of the West. On the other hand, this can be interpreted as a form of self-criticism, a realization that the West has always been violent regardless of Islamic terrorism (“they are bad, but we are no angels either”).

The organization of the broad categories can be viewed along two independent dimensions. First, it is easy to recognize that two broad categories on the left, Enemy and Other-perception, describe “Them” in very general terms. Readers are compelled to learn about them, not just as the enemy, but to clarify who they are and how shall they perceive them. Readers look for a wider context of where the enemy comes from and why. In contrast,

two broad categories on the right describe “Us”. They amount to the construction of a sense of belonging and a safe community, “Our” community. This leads to the introspection of who they are in the face of the enemy. Both “Us” and “Them” can be recognized sharper when placed in opposition to each other. This binary distinction can be interpreted as *Establishing boundaries*: Enemy vs Security, threat/danger vs safety/belonging, unknown vs known, Other vs Self, “Them” vs “Us”.

Second, two broad categories in the bottom amount to *Constructing identities*. The disruptive situation makes the public interested not only in the immediate concerns but in a more comprehensive and abstract perspective on the event. The public becomes curious about Islam not exclusively from a retaliatory perspective. On the other hand, that same public is looking for historic and current facts about Western nations and their political/social organization, neither of which is immediately related to the attacks. Both arising interests point to the need for an identity, i.e., the understanding of how “We” perceive ourselves in face of the “Them”. This elucidates how they establish the boundary between what we they as shared within their community in contrast to whatever is outside of it (and potentially threatening). Interestingly, public also gets curious about Islam’s *presence in the West*—the place where the “They” meet “Us”.

Next, we examine the 15 terrorist attacks through lenses of excess attention Z . For each attack, we calculate its total excess attention by summing over all broad categories. This leads to a single value of Z for each attack, reported in the left panel of Fig. 4. We find the most intense reaction to Paris 2015, Manchester 2017, London Bridge 2017, and Marseille 2017. Other attacks trigger weaker but still noticeable reactions. We could not find meaningful correlations between these values and quantifiable parameters surrounding the attacks, such as the number of deaths or injuries, as it was done by García-Gavilanes et al.³³. The same was found true for the size of the Muslim population in the target countries. Still, we noticed that attacks that took place in the UK and France seem to generate somewhat more excess attention than those occurring elsewhere. A major attack in Brussels in 2016 triggered comparatively weaker interest.

Looking at excess attention of four broad categories concerning the attacks (right panel of Fig. 4), we find that the attention was most intensely diverted to the Enemy category after Paris 2015, and somewhat less so after Manchester 2017 and Marseille 2017. Turku 2017 triggered no interest in the Enemy, while Barcelona 2017 and Liege 2018 exhibit very weak interest in the Enemy. We found the weakest overall reaction in the Security category, where seven attacks did not react at all. In contrast, London Bridge 2017 has an extreme reaction in this category. Redirection to the Other-perception category is somewhat related to that of the Enemy category. Many attacks show similar values for both, with Manchester 2017 and Marseille 2017 being the highest. For the Self-perception category we find moderate values across all attacks, but not much correlated with the Security. The intensities of reactions across four broad categories are unique for each event, the strongest being for Paris 2015.

This paints the picture of each attack having a context of its own and impacting the public in a particular way. The public reaction to attacks is diverse and heterogeneous, with different attacks eliciting varying degrees of response within different broad categories. Nevertheless, we emphasize that despite these nuances, all reactions underscore a division between “Us” and “Them”. Furthermore, we stress that the extent of this dichotomy varies across attacks, although we were unable to identify any external variables that could explain these variances.

Returning the collective attention to ‘normal’. After an attack, how gradually is the public attention restored to routine concerns? Does this depend on the attack and/or the broad category? We look at three examples: Paris 2015, Brussels 2016, and London Bridge 2017. For each of them, we construct a time series that shows

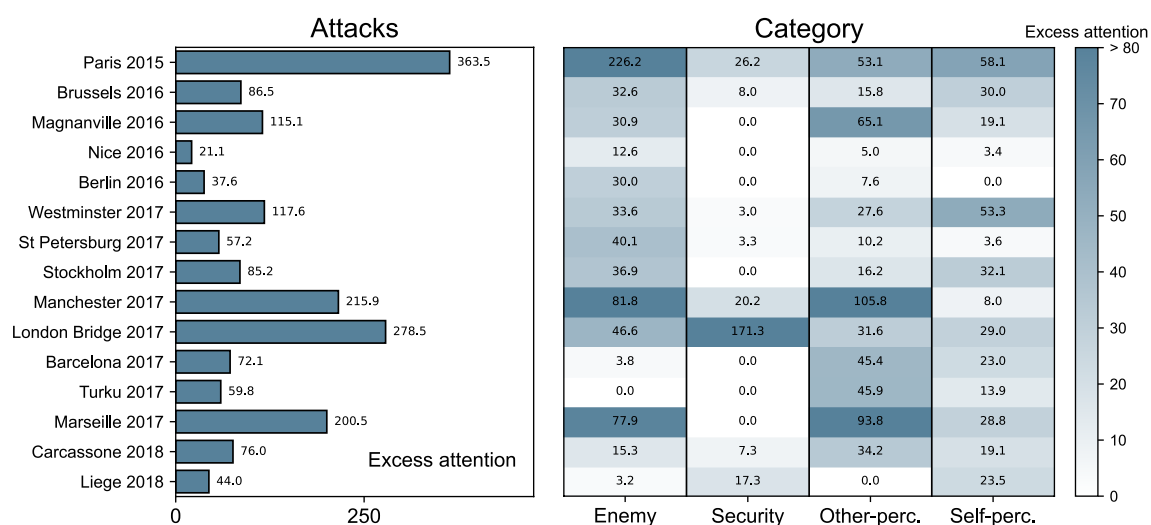


Figure 4. Intensity of the public reaction due to terrorist attacks, which are shown vertically by their abbreviations (“The sample of terrorist attacks” and Appendix A). Left panel: total excess attention for every attack aggregated cumulatively over all of the broad categories. Right panel: total excess attention for every attack broken down over four broad categories, which are shown horizontally. The color bar is truncated for excess attention > 80 for better clarity.

the decay of public attention to each broad category. Results for the period starting a week before the attack to 3 weeks after the attack are shown in Fig. 5.

After every attack, the decrease pattern of collective attention is similar for broad categories of Enemy, Security, and Other-perception, but it is unique to each event. Specifically, for Paris 2015 we find the slowest decrease followed by a sudden increase 3 weeks after the event in all three broad categories. In contrast, the patterns for Brussels 2016 and London Bridge 2017 are very similar: the decrease is relatively constant and quick for Enemy and Security, and somewhat more gradual for Other-perception. The values 3 weeks after these two attacks are similar to those before the attacks. This is not the case for Paris 2015, which again confirms that it was a unique event, with effects persisting long after the event and even peaking again after 3 weeks. In opposition to these three broad categories, we find a different situation for Self-perception. The behavior for Paris 2015 is close to that of its other three broad categories. For Brussels 2016 and London Bridge 2017 we find a rather irregular pattern, characterized by a mix of decay and sudden spikes. These observations suggest that this broad category, pointing to introspection and soul-searching, is the longest-lasting effect of the terrorist trauma.

Discussion and conclusion

Our findings are three-fold. First, the exhaustiveness of Wikipedia enabled us to systematically study how the public prioritizes their interests in the wake of a terrorist attack and how this arrangement of interests differs from the usual trends. Second, we precisely measured which among them receive more and which had less attention. Third, we found consistent patterns in these awoken interests, grounded in the existing ecosystem of beliefs in the affected European societies.

The reaction of the international English-speaking public to the terrorist attacks reveals the deeply rooted binary distinction between East and West, Islam and Europe, “Them” and “Us”. Amid shock, a natural human (individual and collective) reaction is to search for a *community* to belong to. This community has to have a meaning and identity. But the identity is always relational and constructed in opposition to “Others”^{50, 51}, perceived as outsiders or even enemies. The enemy serves as a *mirror* offering a (real or imaginary) distinction between

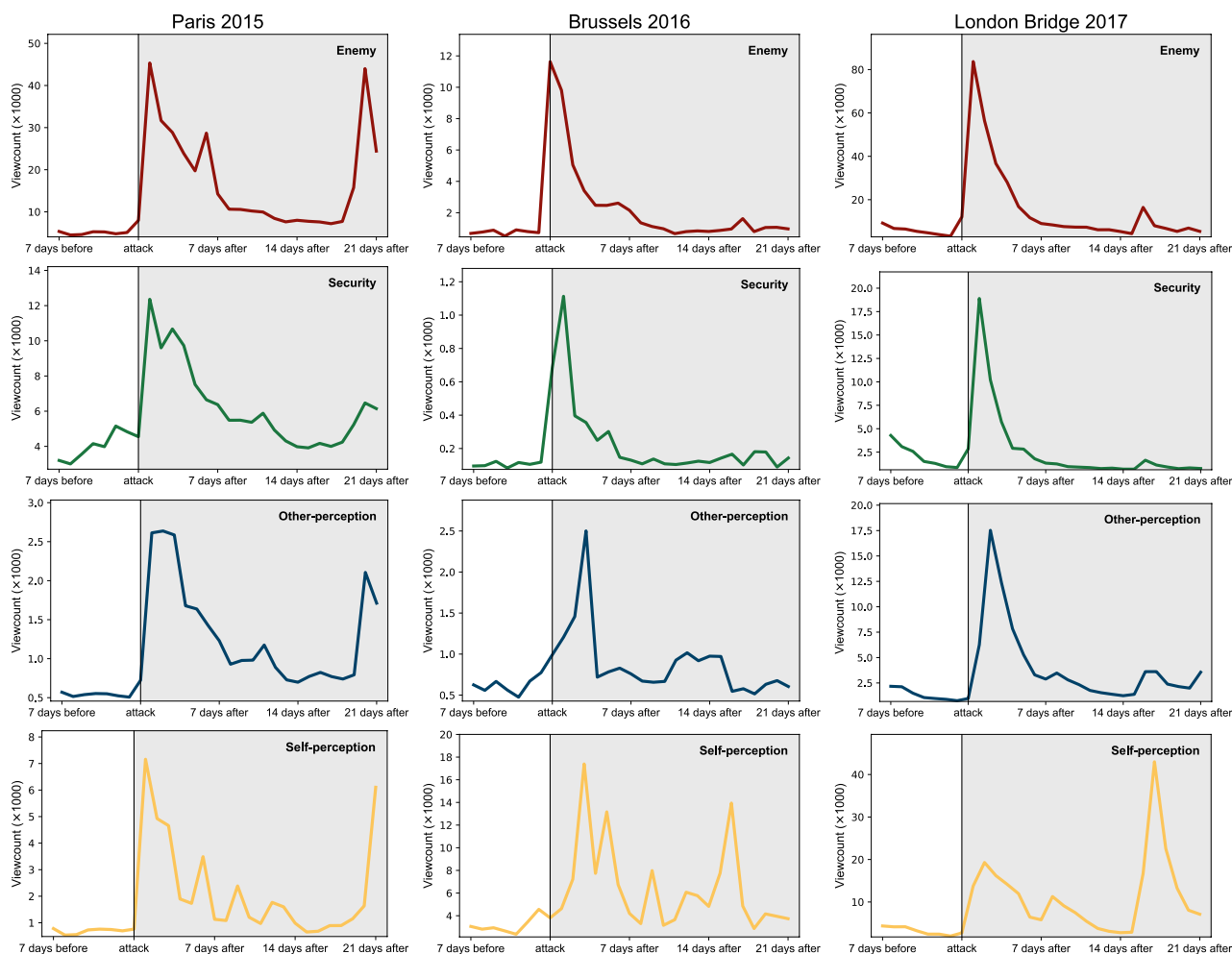


Figure 5. Dynamics of decay of collective public attention after terrorist attacks. Three top row panels: the cumulative time series of viewcounts for Paris 2015, Brussels 2016 and London Bridge 2017, for the broad category Enemy. The time of each attack is marked by a vertical line and the period after it is shaded in grey. Second, third, and fourth rows of panels: the same for the other three broad categories.

“our” group and their group. This makes “Others” an indispensable ingredient for creating feelings of belonging. When this becomes a part of the collective social imaginary, it further fortifies the boundaries between the groups. The construction of social identity presupposes that the in-group differs significantly from the out-group. It also requires an ongoing process of maintaining and legitimizing these boundaries. In this case, “Us” is a transnational society in the making, based on the characteristics, values, heritage, and history of the Anglo-American, European, Western world. The “Western” world has been identified in opposition to Islam through large parts of its history—ranging from the conflicts throughout the Middle Ages, such as the Crusades and the Ottoman expansion, to the more contemporary tensions, such as the support for Israel and the Iraq wars⁴. Terrorist acts—especially if perpetrated by the religious extremists—are ideal mechanisms for recreating this effect.

In conjunction with a community and knowing what is and what is not part of it, we also need an understanding of “Us”, a narrative about *who we are*. This narrative is to contrast the narrative about “Them”, which is to explain *who they are*. Both narratives are usually articulated as a series of attributes about either side, as clearly revealed by the found broad categories. Specifically, the narrative of threatening Enemy to which “we” respond by Security contributes to the symbolic substance of “Us”, who are residing on “our” side of the boundary⁵¹. The social imaginary of the community pertains to a discursive manner of self-understanding, emerging via stories that people tell about themselves and thus construct their collective belonging⁵². In that regard, identities serve as narratives that enable the continuity of a community, form the foundation of the social collective memory, and unify a group through space and time^{53,54}.

The imminent threat may generate other ways of defining the enemy. A prominent human feature is to search for patterns in seemingly chaotic situations while maintaining the feeling of connection. We found this in the interest for *conspiracies*, which replace a visible enemy with a manufactured one. Worrisome adherence to online conspiracy theories is a large topic within computational social science^{55–58}.

The last decades saw individuals’ lives grow apart from large bureaucratic institutions⁵⁹. Individualism and consumerism caused the erosion of the collective loyalties, boosting the scepticism towards the state authority. This trend has been already observed in sociological classics⁶⁰ and accelerated by modern technology and globalization^{61–63}. Interestingly, terrorism seems to push in the opposite direction, towards re-establishing the bond with (presumably) protective institutions, operating at *national* level and provided by state authority.

We also found that intense focus on topics arising from terrorist attacks dissipates after a few weeks. Typically, it is overwhelmed by the never-yielding interest in entertainment. This is hardly surprising, given that the public is known to shorten its attention span^{43,44}, even during the COVID crisis⁴⁵. This is in agreement with one of our results, namely, that the public tends to collectively remember only the most recent attacks (cf. Fig 2). In Appendix C, we show a side result: The amount of collective attention redirected from other routine interests due to terrorism is insignificant from the perspective of overall viewing behavior on Wikipedia (about 200–300 million views per day with weekly and seasonal oscillations). In other words, none of the attacks was ‘that interesting’ to divert major and lasting attention from entertainment.

Individuals’ prioritization of topics expresses narratives that they were brought up with. These involve social memories as a basis of collective (for instance national) identity, which includes cultural norms, notions of authenticity, power relations, etc. Our findings indicate that for international English-speaking audiences these narratives are centered around security, enemy, self-perception, and other-perception. In them we recognize a *search for belonging and meaning* in a globalizing late modern society, triggered by a sudden disruption. These results are in accord with the classical studies on the relationship between Europe (the West) and its neighbours (the East), referring especially to the collective perceptions of the “significant other”⁴. They call for reconsidering the classical ideas of nationalism, thus advancing the constructivist understanding of society⁵¹, which is detached from the actual physical places⁶⁴. In that regard, we are consistent with the contemporary scholarly consideration of belonging, which is increasingly shifting towards transnational and virtual social reality, articulated through imagined communities and forged via “fragile communication bonds” such as common usage of English Wikipedia^{65–67}.

Our results could benefit from an independent verification via sources other than Wikipedia, primarily involving social media. Still, the popularity of any tweet could be due to lack of a systematic repertoire of tweets, rather than genuine public interest in that tweet’s content. On the other hand, what else, besides terrorist attacks, diverts attention to the 69 articles from Table 1? Perhaps some articles react *only* to the terrorist attacks and could serve as ‘markers’. We examined this in the Appendix D and found no evidence for them. Another methodological challenge is the robustness against the modifications of exclusion criteria. We check this, and found that key results are robust against all meaningful modifications. For more on this and other robustness checks see Appendix E.

There are two main directions for advancing the presented work. First, spontaneous reaction to disruptive situations naturally involves languages other than English, both in Wikipedia and elsewhere. This makes online sources in other languages prime target of future work. Second, from how we made our sample of attacks, it just so turned out that all 15 of them were committed by Islamic extremists. But what about terrorist acts perpetrated, for example, by Western far-right groups? We fully acknowledge that terrorism can have very diverse origins and that public reactions to it must be investigated. In this light, it is also of interest to look at a wider time span, starting for example from Twin Towers attack in 2001.

Our conclusion is that while unfortunately terrorism is not likely to vanish soon, studies like this one can help understand the intricacies of the public reactions to it, and minimize the consequential societal damages such as extremism and polarization.

Data and methods

The sample of terrorist attacks. We considered all confirmed acts of terrorism in Europe starting from November 2015 Paris attacks until the end of 2018. Using Wikipedia's information, we found 43 such acts but kept only those with at least two casualties excluding the attackers. This led us to a sample of 15 terrorist attacks across Europe (see Fig. 6 for the geographical location and the timeline of the incidents). Note that the available Wikipedia viewership logs commence from July 2015. This restricts our options in selecting the attacks and makes Paris 2015 the most suitable event to start from. Of course, had we had the access to data prior to 2015, we would have started our analysis from the New York attack of 2001 and possibly broaden it beyond Europe.

Data preparation. Our overarching aim is to examine the visits to *all* articles in English Wikipedia during the relevant period and identify those articles that are visited more frequently (receive excess attention) after a terrorist attack. We started by establishing a list of entries that can be considered valid articles^{68,69}. As of August 2019, this list contained 14,733,158 entries. Through Wikimedia API⁷⁰ we can obtain the exact viewcount (number of visits) for an article on any day. This *daily viewcount* is a great proxy for the attention devoted to that topic by the international English-speaking audience. We retrieved the time series of daily viewcounts for the entire list with 14,733,158 entries from 1 January 2015 to 8 August 2019 (1680 days). Each value is the total number of visits to an article on a given day, cumulatively generated by entire online traffic from all devices with access to the Internet. We want to examine only those articles whose viewcounts spiked after terrorist attack(s) and whose content can be meaningfully connected to the attack. After the extensive exclusion procedure described below, we identified 69 such articles.

Sampling the articles around the time of Paris 2015 attack. We sample the articles for Fig. 1 by considering the 12 days surrounding the event. We found this time window most suitable for illustrating the dynamics of collective attention around the attack. We looked at the top 5 articles by viewcount on each of these 12 days. We then selected articles, that on at least 2 days (of these 12 days) received more than 100,000 views. This gave us the sample of 17 articles, shown vertically in Fig. 1.

Definition of excess attention. Excess attention, Z , is defined as follows. Let A be a Wikipedia article and T the time of a specific attack. The mean $\mu(A, T)$ and standard deviation $\sigma(A, T)$ of viewcounts of A before the attack are calculated starting from 1st of January 2015 as:

$$\mu(A, T) = \frac{1}{T-1} \sum_{t=1}^{T-1} w(A, t) \quad , \quad \sigma(A, T) = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T-1} (w(A, t) - \mu(A, T))^2} \quad ,$$

where $w(A, t)$ is the viewcount of A on the day t . Next, we compute $v(A, T)$, the average viewcount of A in the week following the attack:

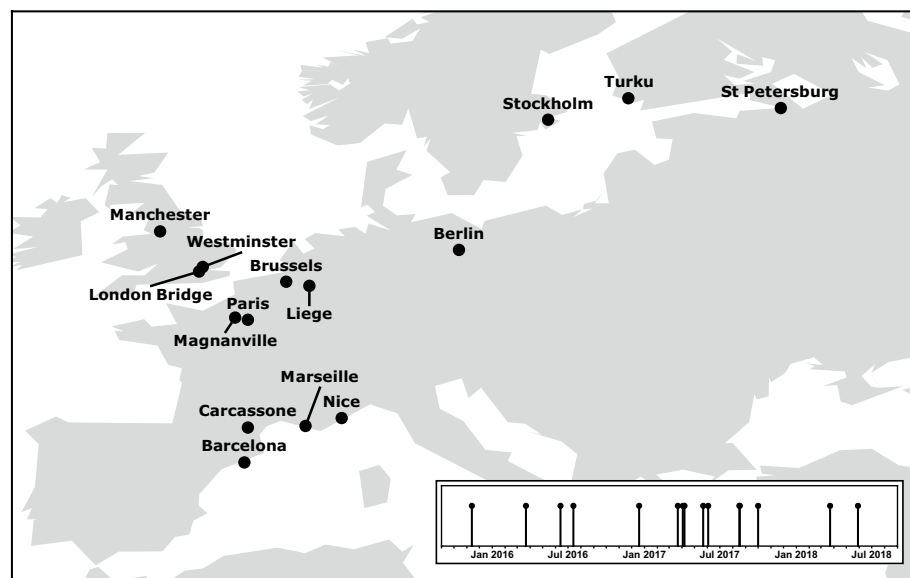


Figure 6. Locations (cities) of 15 terrorist attacks comprising our sample. Attacks occurred between November 2015 (Paris) and May 2018 (Liege), as illustrated in the inset. See Appendix A for a comprehensive overview and details of each attack.

$$v(A, T) = \frac{1}{7} (w(A, T) + w(A, T + 1) + \dots + w(A, T + 6)).$$

Our choice of a week (7 days) as the period for averaging is motivated by pronounced weekly oscillations of overall viewcounts and in line with³³ (see also Appendix C). Now we define $Z(A, T)$ as:

$$Z(A, T) = \frac{v(A, T) - \mu(A, T)}{\sigma(A, T)}. \quad (1)$$

Excess attention measures the increase of mean viewcounts in a week after the event compared to the mean viewcount before the attack, expressed in the units of standard deviation characterizing the trend prior (our measure of Excess Attention is conceptually similar to Z-score, even if it is not exactly the same.) It can be computed for any article A in association with any attack T . For example, if an article displays $Z = 3$ in association with T , it means that in the week following T the visits to A have increased by 3 standard deviations compared to the average before T . This is our main tool in identifying the articles that have reacted to an attack in contrast to those that did not.

For Fig. 2 (right panel), since we need a more stable estimates of μ and σ , we exclude the first 2 weeks after the older incident. For Fig. 4 (left panel), we assigned the total (cumulative) Z to each attack by summing Z s for all articles that reacted to that attack, cumulatively across all broad categories. We did the same for Fig. 4 (right panel), but summarizing each broad category separately.

Article exclusion steps. Step 1: Removal of stubs and redirects. Wikipedia uses the term “stub” for abnormally incomplete articles. Using the stub tags we identified and removed them. We also removed articles that only carry an alternative title and a link to the original article with the more conventional title (“redirects”). Starting from the complete list of articles in English Wikipedia, excluding all stubs and redirects left us with 5,876,878 articles.

Step 2: Exclusion of articles with low viewership. To filter our low view articles, we used a cut-off based on the maximal daily viewcount. We kept only the articles whose daily viewcount at least once reached 300. See Fig. 7 for the distribution of maximum daily views and our cut-off determined based on the “elbow rule”. This step left us with 912,192 articles.

Step 3: Removal of articles with small excess attention after attacks. Next, for each attack separately, we exclude articles that received no marked increase in attention immediately after the incident (true negatives). As threshold we use $Z = 3$: Assuming that viewcounts are normally distributed, the chance of observing $Z \geq 3$ at random is $p \sim 0.001$ (see also Appendix E). Upon excluding articles with $Z < 3$, we were left with 9000–24,000 articles per attack. To better illustrate this step, we show in Fig. 8 the viewcounts for three example articles around the time of Paris 2015. This filter is successful in detecting articles whose increase in viewcounts is related to the attack (true positive), but unsuccessful in filtering out articles that had an increase in viewcounts due to other unrelated events (false positive). An example is G20 that received spike in attention due to G20 meeting that accidentally took place around the same time as Paris 2015. We deal more with false positives later.

Step 4: Automated removal based on content Inspecting the articles which passed the previous filters, we noticed common patterns in their topics, pointing to an overarching picture of the arising interests. Still, as too many events occur simultaneously, there were still too many false positives. To remove them, we employed semi-supervised Google’s text-miner BERT (Bidirectional Encoder Representations from Transformers)^{71,72}. It outputs a score that expresses how similar is the content of some text to the content of the training text. We prepared the training text by taking a random sample of 10% of articles for each attack and manually filtering out the articles

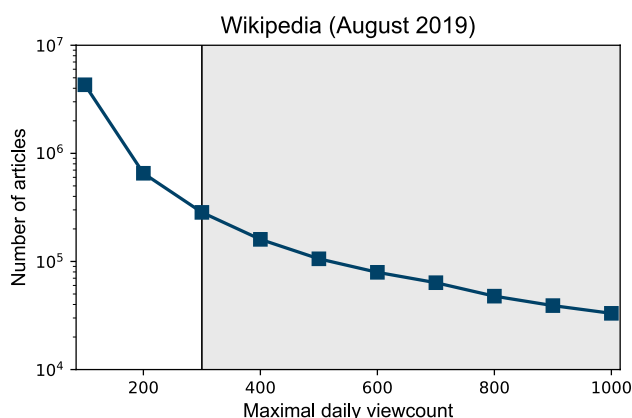


Figure 7. The distribution of maximal daily viewcounts (maximal number of views reached on a single day) between 1 January 2015 and 8 August 2019, for all 5,876,878 articles after Step 1. We kept only the articles in the shaded area, i.e., excluded those that were never visited more than 300 times on a single day. Note a slight change of slope near 300, making this a natural choice for cut-off.

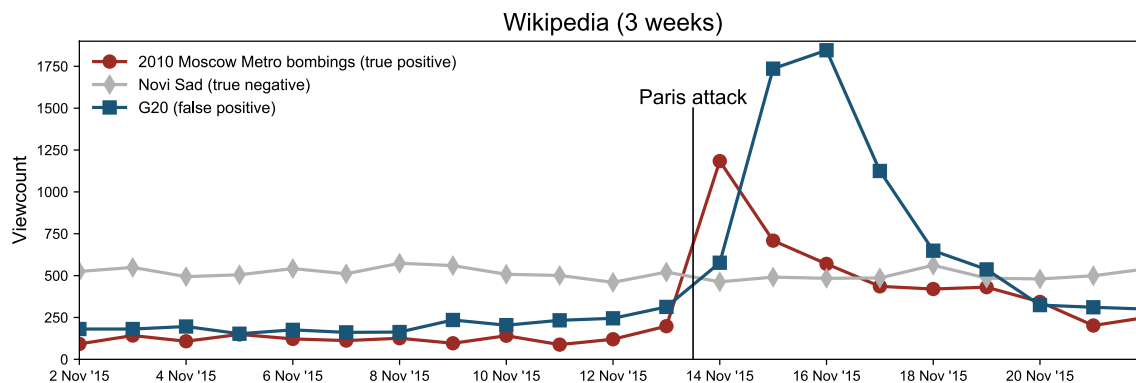


Figure 8. Viewcounts of three articles around the time of the November 2015 Paris attack. Attention on the article in red peaks immediately after the attack, and given its content, we reasonably expect that this was a reaction to the attack (true positive). The article in grey exhibits no qualitative change, and indeed, given its topic (a city in Serbia), we did not expect any (true negative). The attention is devoted to the example in blue peaks, but this article's content precludes us to attribute this to the attack (false positive).

which were not conceptually related to the attacks. We then extracted the summaries (Wikipedia calls 'summary' a brief description of the article's content, available immediately after the title.) of all remaining articles and trained BERT on them. Then, we had BERT recognize the similar content in the summaries of all articles surviving Step 3. Upon completion, we were left with 1000–8000 articles per attack. We evaluated BERT's performance by manually inspecting a 10% sample of thus excluded articles and indeed found them all true negatives.

Step 5: Distilling the reactions Our next aim is to identify themes that systematically appear in public interests and are not connected to a single attack. We put the surviving articles back together and remove those that reacted to less than three independent attacks, not counting as independent the reactions to events in the same week. In establishing what constitutes a reaction to an attack, we made two additional requirements:

1. Looking at some obvious false positives, we noticed that their viewcounts start increasing before the attack and hence clearly *not due to* the attack. To account for this we introduce $Z_0(A, T) = Z(A, T - 7)$, the Z -score for a week immediately before the attack. We found Z_0 to be unusually large for most false positives. Yet, a slightly increased Z_0 is not problematic as long as it is followed by a considerably higher Z post-attack. For consistency with earlier steps, we requested

$$Z(A, T) > 2 + \frac{3}{3 - Z_0(A, T)}$$

for every reaction of A to T . Hence, for an article with $Z_0 = 1$ we request at least $Z = 3.5$, for articles with $Z_0 = 2$ at least $Z = 5$, etc.

2. Since the mean $\mu(A, T)$ appears in the denominator of Eq. (1), Z is more volatile when $\mu(A, T)$ is small. An unusually high Z found for A with small μ could easily be an artifact (division by a number very close to zero). To account for this, we excluded the reactions of articles A to attacks T whenever we found $\mu(A, T) < 20$.

See Fig. 9 for illustration. This step left us with 86 articles.

Step 6: Qualitative (manual) exclusion based on content. We next inspected a broader context of every remaining article around the time of each observed reaction. We assigned a thematic code to each article using open coding⁷³ based on two criteria:

- consideration of the articles' content and its thematic relevance from the perspective of each particular attack,
- timing of its appearance in terms of accidental coinciding with other unrelated but simultaneously occurring events (possible false positives).

We excluded 17 articles as we found them to coincide with: sporting events; launching of movies and TV shows; music events, personalities and their media coverage; and other accidentally coinciding events (solar eclipse, legislation procedures, cultural events, etc.). There were three articles on the Syrian civil war: the Palmyra offensive in March 2016, the Russian offensive in November 2015, and the US response to the alleged chemical attack by the Syrian regime in April 2017. Upon very close inspection we could not connect them to the terrorist events in Europe. This left us with 69 articles, shown in Table 1.

Construction of specific categories. Table 1 is a succinct list of what the public becomes interested in following a terrorist attack. Not surprisingly, it includes terrorism, weapons, and anti-Islamism, but also national identity, the violent history of the West, and conspiracy theories. As many articles have similar content, we used

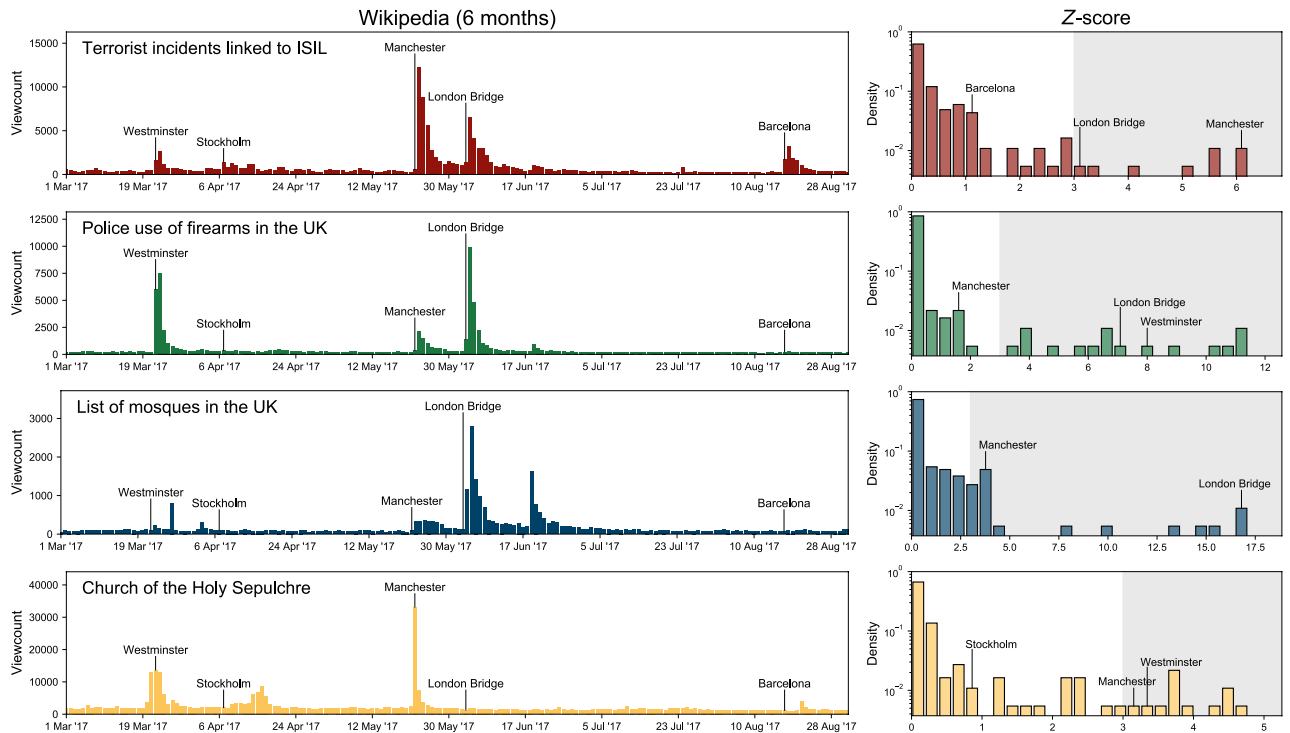


Figure 9. Examples of four articles identified in Step 5. Top left: viewcount time series of *Terrorist incidents linked to ISIL* from 1 March 2017 to 1 September 2017. Attacks that occurred during those 6 months are indicated. Top right: the histogram of Z-scores for that article for the entire time span (region $Z > 3$ is shaded). Other panels: same as on the top panels (left and right) for other three examples of articles. These plots demonstrate how well the threshold $Z = 3$ separates day-to-day fluctuations from what we can consider a reaction to an attack.

axial coding⁷³ to extract their thematic codes and connect them. By this, we manually cluster the articles into meaningful groups, defined by the coherency of their content. We call these groups *specific categories* and assign a name to each of them that summarizes the content. The 69 articles from Table 1 are arranged into the following 19 specific categories.

1. **Terrorism in general.** Definitional aspects of terrorism, such as what constitutes a mass shooting and what is a massacre (3 articles).
2. **Islamic terrorism.** Past Islamic terrorist acts and organizations, regardless of location, context, and targets (9 articles).
3. **Non-Islamic terrorism.** Far-right and far-left terrorism and extremism in the West, not related to Islam (3 articles).
4. **Terrorism in Western countries.** Summaries of terrorist activities in the West, including those not part of our sample, regardless of religious/ideological origin (7 articles).
5. **Conspiracies.** Possible conspiracies behind or related to the attacks (1 article).
6. **National security.** Western security and defence institutions, their official weaponry, conscription policies and related legal frameworks (9 articles).
7. **Weapons.** Weapons for personal defence and comparisons among them (e.g. AK47 vs. M16), firearms legislation and purchasing regulations (7 articles).
8. **Transport options.** Public transportation infrastructure, including air travel, railroads, and highways (5 articles).
9. **Insurance options.** Financial security and purchasing of insurance (2 articles).
10. **Equipment for survival.** Purchasing equipment for surviving in the wild (1 article).
11. **Islam in general.** Islam religious figures (1 article).
12. **Islam in Western countries.** Muslim religious institutions, figures, and practices in the West (4 articles).
13. **Islam in Non-western countries.** Role and history of Islam in non-Western countries (1 article).
14. **About Middle East.** Locations and history of the Middle East (1 article).
15. **National identity.** Generalities about Western nations and their political order (1 article).
16. **Religious identity.** Christian religious institutions in the West (1 article).
17. **Activism in general.** Western political activists and writers, not related to Islam (2 articles).

Andrew O'Hagan	Kyle Kulinski	Devin Nunes
James Wickstrom	Paul Weston (politician)	Sebastian Gorka
Tommy Robinson (activist)	False flag	Orvis
TD Banknorth	Thornburg Mortgage	Moinuddin Chishti
Islam in Qatar	Ed Husain	Islam in Poland
List of mosques in the United Kingdom	Quilliam (think tank)	Susa
Juan Carlos I	Conscription in the United States	Crisis actor
Kimberly Dozier	List of police firearms in the United Kingdom	Police use of firearms in the United Kingdom
Ring of steel (London)	Spanish Armed Forces	UK Threat Levels
USS Cheyenne (SSN-773)	Church of the Holy Sepulchre	Lone wolf (terrorism)
Mass shooting	Massacre	Bhagwan Shree Rajneesh
List of terrorist incidents in Australia	List of terrorist incidents in Great Britain	List of terrorist incidents in London
Terrorism in Europe	Terrorism in Greece	Terrorism in Sweden
2002 Los Angeles International Airport shooting	2004 SuperFerry 14 bombing	2007 Glasgow Airport attack
2010 Moscow Metro bombings	Abduction of Russian diplomats in Iraq	List of terrorist incidents linked to ISIL
Muslim Brotherhood	Rick Leventhal	Terrorism in the Philippines
Basque conflict	Basque National Liberation Movement	Timeline of Real Irish Republican Army actions
Big Sky Airlines	Boston-Maine Airways	Corporate Air
Dakota, Minnesota and Eastern Railroad	Eclipse Aviation	Charles B. McVay III
Grdelica train bombing	Greek War of Independence	Legitimacy of the NATO bombing of Yugoslavia
List of events named massacres	The Man with the Iron Heart	Assault weapon
Comparison of the AK-47 and M16	Gun culture in the United States	Gun politics in the United States
Jews for the Preservation of Firearms Ownership	United States v. Miller	Universal background check

Table 1. The list of 69 English Wikipedia articles that have reacted to at least three independent attacks. They jointly represent the content to which terrorist acts divert public attention (irrespective of the particular attack's details).

18. *Anti-Islam activism*. Western far-right politicians, public figures and anti-Islam extremists (5 articles).
19. *Violent history of Western countries* Violent historic events in the West, famous battles in the two World Wars, local wars of independence (6 articles).

Arrangement of individual articles from Table 1 into above 19 specific categories is shown in form of table in Appendix B.

Construction of broad categories. The above organization of articles into specific categories is not immune to ambiguities, such as placing an article into one category or another or merging or splitting categories. We resolve this by noticing similarities across specific categories' thematic codes. For instance, categories 1–4 revolve around terrorism, whereas categories 11–14 deal with varying aspects of Islam and its relation to (real or perceived) perpetrators. This allows us to coarse-grain 19 specific categories into four *broad categories*: Enemy (Specific categories 1–5, 23 articles), Security (Specific categories 6–10, 24 articles), Other-perception (Specific categories 11–14, 7 articles), and Self-perception (Specific categories 15–19, 15 articles).

We verified that the topical coherency of broad categories is robust against the above ambiguities and even against the most meaningful modifications of the exclusion choices (e.g. requesting the reaction to only two independent attacks). To further increase the reliability and validity, we sought to define their thematic codes in the context of discourse analysis. We applied intra-coding in terms of re-checking broad categories' thematic codes 12 months after the initial coding process and found them to be consistent (We decided against deploying inter-coded reliability since we did not find it compatible with recursive and incremental nature of our coding process.). Thus constructed broad categories are *universal* and interpretable as the broad public interests awakened by any terrorist act.

Data availability

The datasets behind this study are freely available from Wikipedia. For easier access by the scientific community, we deposited them at ZENODO repository, at <https://zenodo.org/record/7198794#.Y0mCXUpBxkg>. We offer our consultation to whoever wishes to work on this data.

Received: 14 September 2022; Accepted: 19 July 2023

Published online: 01 August 2023

References

- Magdy, W., Darwish, K., & Abokhodair, N. Quantifying public response towards Islam on Twitter after Paris attacks. (2015). [arXiv:1512.04570](https://arxiv.org/abs/1512.04570).
- Strebel, M. A., & Steenbergen, M. R. The impact of the November 2015 terrorist attacks in Paris on public opinion: A natural experiment. In *Department of Political Science, University of Zurich* (2017).
- Vasilopoulos, P. *et al.* Fear, anger, and voting for the far right: Evidence from the November 13, 2015 Paris terror attacks. *Polit. Psychol.* **40.4**, 679–704 (2019).
- Said, E. *Orientalism: Western Concepts of the Orient* (Pantheon, 1978).
- Rane, H., Ewart, J. & Martinkus, J. *Media Framing of the Muslim World* (Palgrave Macmillan, 2014).
- Kearns, E., Betus, A. & Lemieux, A. Why do some terrorist attacks receive more media attention. *Than Others* **366**, 985–1022 (2019).
- Slone, M. Responses to media coverage of terrorism. *J. Conflict Resolut.* **44.4**, 508–522 (2000).
- Pfefferbaum, B. *et al.* Disaster media coverage and psychological outcomes: Descriptive findings in the extant research. *Curr. Psychiatry Rep.* **16.9**, 464 (2014).
- Iyer, A. *et al.* Understanding the power of the picture: The effect of image content on emotional and political responses to terrorism. *J. Appl. Soc. Psychol.* **44**, 511–521 (2014).
- Conway, M. Determining the role of the internet in violent extremism and terrorism: Six suggestions for progressing research. *Stud. Conflict Terrorism* **40.1**, 77–98 (2017).
- Hopwood, T. L. & Schutte, N. S. Psychological outcomes in reaction to media exposure to disasters and large-scale violence: A meta-analysis. *Psychol. Violence* **7.2**, 316–327 (2017).
- Monfort, E. & Afzali, M. H. Traumatic stress symptoms after the November 13th, terrorist attacks among young adults: The relation to media and emotion regulation. *Compr. Psychiatry* **752017**(68–74), x (2019).
- Thompson, R. R. *et al.* Media exposure to mass violence events can fuel a cycle of distress. *Sci. Adv.* **54**, eaav3502 (2019).
- Garcia, D. & Rimé, B. Collective emotions and social resilience in the digital traces after a terrorist attack. *Psychol. Sci.* **30**, 617–628 (2019).
- Schuurman, B. Research on terrorism, 2007–2016: A review of data, methods, and authorship. *Terrorism Polit. Violence* **32**(5), 1011–1026 (2020).
- Álvarez-Benjumea, A. & Winter, F. The breakdown of antiracist norms: A natural experiment on hate speech after terrorist attacks. *Proc. Natl. Acad. Sci.* **117**(37), 22800–22804 (2020).
- Python, A. *et al.* Predicting non-state terrorism worldwide. *Sci. Adv.* **7**, 31 (2021).
- Giani, M., Epifanio, M., & Ivandic, R. Wait and see? Public opinion dynamics after terrorist attacks. (2021). <https://doi.org/10.31235/osf.io/qt7s4>.
- Yasseri, T. & Bright, J. Wikipedia traffic data and electoral prediction: Towards theoretically informed models. *EPJ Data Sci.* **5**(1), 1–15 (2016).
- Berrebi, C. & Klor, E. F. On terrorism and electoral outcomes: Theory and evidence from the Israeli–Palestinian conflict. *J. Conflict Resolut.* **506**, 20 (2006).
- Bali, V. A. Terror and elections: Lessons from Spain. *Electoral Stud.* **26**, 3 (2007).
- Berrebi, C. & Klor, E. F. Are voters sensitive to terrorism? Direct evidence from the Israeli electorate. *Am. Polit. Sci. Rev.* **50**, 6 (2008).
- Gould, E. D. & Klor, E. F. Does terrorism work?. *Q. J. Econ.* **125**, 4 (2008).
- Abrahms, M. The political effectiveness of terrorism revisited. *Comp. Polit. Stud.* **45**, 3 (2012).
- Frey, B. S., Luechinger, S. & Stutzer, A. Calculating tragedy: Assessing the costs of terrorism. *J. Econ. Surv.* **21**, 1 (2007).
- Benmelech, E., Berrebi, C. & Klor, E. F. The economic cost of harboring terrorism. *J. Conflict Resolut.* **54**, 2 (2010).
- Becker, G. S., & Rubinstein, Y. Fear and the response to terrorism: An economic analysis. In *Center for Economic Performance* (2011).
- Benmelech, E., Berrebi, C. & Klor, E. F. Economic conditions and the quality of suicide terrorism. *J. Polit.* **74**, 1 (2012).
- Huff, C. & Kertzer, J. D. How the public defines terrorism. *Am. J. Polit. Sci.* **62**, 1 (2017).
- Bove, V. *et al.* Are the effects of terrorism short-lived? (Tech. rep, World Institute for Development Economic Research (UNU-WIDER), 2022).
- Ratkiewicz, J., Flammini, A., & Menczer, F. Traffic in social media I: Paths through information networks. In *2010 IEEE Second International Conference on Social Computing*, 452–458 (2010).
- Yoshida, M., *et al.* Wikipedia page view reflects web search trend. In *Proceedings of the ACM Web Science Conference*, 1–2. <https://doi.org/10.1145/2786451.2786495> (2015).
- García-Gavilanes, R., Tsvetkova, M. & Yasseri, T. Dynamics and biases of online attention: The case of aircraft crashes. *R. Soc. Open Sci.* **3**(10), 160460 (2016).
- Mestyán, M., Yasseri, T. & Kertész, J. Early prediction of movie box office success based on Wikipedia activity big data. *PLoS One* **8**(8), e71226 (2013).
- Yasseri, T. & Bright, J. Can electoral popularity be predicted using socially generated big data?. *IT Inf. Technol.* **56.5**, 246–253 (2014).
- Ferron, M. & Massa, P. Studying collective memories in Wikipedia. *J. Soc. Theory* **3**(4), 449–466 (2011).
- Helen, S. M. *et al.* Quantifying Wikipedia usage patterns before stock market moves. *Sci. Rep.* **31**, 1–5 (2013).
- Kanhubua, N., Nguyen, T. N., & Niederée, C. What triggers human remembering of events?: A large-scale analysis of catalysts for collective memory in Wikipedia, 341–350 (2014).
- McIver, D. J. & Brownstein, J. S. Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time. *PLoS Comput. Biol.* **10.4**, e1003581 (2014).
- Thalhammer, A., & Rettinger, A. PageRank on Wikipedia: Towards general importance scores for entities. In *European Semantic Web Conference*, 227–240 (Springer, 2016).
- Ban, K., Perc, M. & Levnaji, Z. Robust clustering of languages across Wikipedia growth. *R. Soc. Open Sci.* **41.0**, 171–217 (2017).
- García-Gavilanes, R. *et al.* The memory remains: Understanding collective memory in the digital age. *Sci. Adv.* **3**(4), e1602368 (2017).
- Candia, C. *et al.* The universal decay of collective memory and attention. *Nat. Hum. Behav.* **3**, 82–91 (2019).
- Lorenz-Spreen, P. *et al.* Accelerating dynamics of collective attention. *Nat. Commun.* **10**(1), 1–9 (2019).
- Ribeiro, M. H. *et al.* Sudden attention shifts on Wikipedia during the COVID-19 crisis. *Proc. Int. AAAI Conf. Web Soc. Media* **15**, 208–219 (2021).
- Kobayashi, R., *et al.* Modeling collective anticipation and response on Wikipedia. [arXiv:2105.10900](https://arxiv.org/abs/2105.10900) (2021).
- Arnold, M. V. *et al.* Hurricanes and hashtags: Characterizing online collective attention for natural disasters. *PLoS One* **16**(5), 1–20 (2021).
- Kerby, A. P. *Narrative and the Self* (Indiana University Press, 1991).
- Nelson, K. Self and social functions: Individual autobiographical memory and collective narrative. *Memory* **11**, 125–136 (2003).
- Jenkins, R. *Social Identity* (Routledge, 1996).

51. Cohen, A. P. *Symbolic Construction of Community* (Routledge, 2013).
52. Potter, J. *Representing Reality: Discourse, Rhetoric and Social Construction* (Sage, 1996).
53. Halbwachs, M. *On Collective Memory* (University of Chicago Press, 1992).
54. Lowenthal, D. *The Past is the Foreign Country* (Cambridge University Press, 1995).
55. Del Vicario, M. *et al.* The spreading of misinformation online. *Proc. Natl. Acad. Sci.* **113**(3), 554–559 (2016).
56. Van Prooijen, J. W., Douglas, K. M. & De Inocencio, C. Connecting the dots: Illusory pattern perception predicts belief in conspiracies and the supernatural. *Eur. J. Soc. Psychol.* **48**(3), 320–335 (2018).
57. Friedman, R. A. Why humans are vulnerable to conspiracy theories. *Psychiatr. Serv.* **721**, 3–4 (2021).
58. Jake Hofman, M. *et al.* Integrating explanation and prediction in computational social science. *Nature* **595**, 25 (2021).
59. Berger, P. & Luckmann, T. *Modernity, Pluralism and the Crisis of Meaning* (Bertelsmann Foundation Pubs, 1995).
60. Durkheim, É., Suicide, A., Spaulding, J. & Simpson, G. *Study in Sociology* (Routledge, 1951).
61. Castells, M. *The Information Age* Vol. 98 (Blackwell Publishers, 1996).
62. Beck, U. & Elisabeth, B.-G. *Individualization: Institutionalized Individualism and Its Social and Political Consequences* (SAGE Publications Ltd, 2002).
63. Urry, J. Global complexities. *Int. Stud. Rev.* **5**, 2 (2003).
64. Appadurai, A. *Modernity at Large: Cultural Dimensions of Globalization* Vol. 1 (U of Minnesota Press, 1996).
65. Anderson, B. *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (Verso books, 2006).
66. Barth, F. *Ethnic Groups and Boundaries: The Social Organization of Culture Difference* (Waveland Press, 1998).
67. Delanty, G. *Community* (Routledge, 2018).
68. Wikipedia. Wikipedia dumps. <https://dumps.wikimedia.org/enwiki/latest/>. Accessed 09 Sep 2018.
69. Wikipedia. Index of page view statistics for (2015). <https://dumps.wikimedia.org/other/pagecounts-raw/2015/>. Accessed 08 Aug 2019.
70. Wikipedia. Wikimedia REST API. https://wikimedia.org/api/rest_v1/#/. Accessed 08 Aug 2019.
71. Devlin, J. *et al.* Bert: Pre-training of deep bidirectional transformers for language understanding. (2018). [arXiv:1810.04805](https://arxiv.org/abs/1810.04805).
72. BERT. bert-as-service. <https://bert-as-service.readthedocs.io/en/latest/>. Accessed 05 May 2020.
73. Flick, U. *Introducing Research Methodology: Thinking Your Way Through Your Research Project* (Sage, 2020).

Acknowledgements

We acknowledge the support of the Ministry of Education, Science and Technological Development of the Republic of Serbia (Grant No. 451-03-68/2022-14/200125) and by the Slovenian Research Agency (ARRS) via programs P1-0383 and P5-0168, and via project J5-8236. In memory of our dear colleague Novica Bjelica, who helped us set up the datasets.

Author contributions

Z.L., T.Y., S.S., D.B.K. conceptualized the research. M.J., L.S., D.B.K. extracted, pre-processed, and curated the data. M.J., L.S. analyzed the data. Z.L., M.M., T.G. interpreted the results. M.J., L.S. prepared the figures and tables. Z.L., S.S., D.B.K., M.M., T.G., T.Y. wrote the manuscript. All authors reviewed it.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material (Appendices) available at <https://doi.org/10.1038/s41598-023-39035-3>.

Correspondence and requests for materials should be addressed to Z.L.

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.



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

© The Author(s) 2023