Global Terrorism and (De-)Mobilization: Do ISIS’s Deadly Attacks Demotivate, Deter, or Mobilize Supporters?

Joan Barceló * Elena Labzina †

January 1, 2017

Abstract

Although existing research argues that terrorist organizations seek to achieve two goals with their attacks, mobilization of supporters and infusion of terror among the population, the actual empirical evidence on either of them is, at best, scarce. Based on the number of followers of ISIS-related Twitter accounts, we establish a formal strategy to identify a causal impact of terrorist attacks on movements across the online insider-outsider boundary of this extremist organization. This paper employs a novel dataset with 300,842 observations of 13,321 unique Twitter accounts linked to the Islamic State of Syria and Iraq (ISIS), collected during a 127-day period (from March to July 2016). By exploiting the exogenous timing of terrorist attacks as a natural experiment, we find that the number of followers of ISIS-related Twitter accounts significantly reduces in the aftermath of the terrorist attacks. Furthermore, we evaluate the impact of attacks by implementing the interrupted time series analysis on two major attacks: the Brussels bombings in Belgium on March 22, 2016, and the attack in Nice, France on July 14, 2016. Consistently, we find that the number of the followers of the ISIS-related accounts decreases significantly on the days following major terrorist attacks. This observation provides the evidence that ISIS’s terrorist attacks reduce the online audience of their organization. In addition, we differentiate between two mechanisms: demotivation – a change in supporters’ beliefs – and deterrence – online de-mobilization due to fear. To disentangle between the two, we derive a key empirical observation from the deterrence mechanism and examine it empirically. We do not find support for it, and hence conclude that the demotivation effect dominates our main result. Finally, we conclude that the shrunk of ISIS’ Twitter audience may be their sacrifice to accrue benefits in other areas of interest, such as internal mobilization or persuading governments to achieve policy concessions.

Keywords: Terrorism, Violence, Mobilization, Deterrence, Islamic State, Twitter

*Ph.D. Candidate. Departments of Political Science. Washington University in St. Louis. joan-barcelosoler@wustl.edu.
†Ph.D. Candidate. Departments of Political Science and Mathematics. Washington University in St. Louis. elena.labzina@wustl.edu.
## Contents

1  The Impact of Terrorist Attacks on Twitter Followers  
   1.1  Terrorist Attacks as Strategies of Mobilization and Discouragement  ...  8  
   1.2  The Islamic State on Twitter  .............................................. 10  

2  Formal modelling of the response to an ISIS attack  
   2.1  Individual transition model  .............................................. 16  
   2.2  Group comparative statistics  ............................................ 17  

3  Data  
   3.1  Information on ISIS-linked Accounts  ................................. 18  
   3.2  The Cumulative Lagged Measure of Terrorist Attacks  .......... 20  

4  Empirical Strategy  

5  Results  
   5.1  Mobilization or Discouragement?  .................................... 24  
       5.1.1  Case studies: An Interrupted Time Series Analysis of Two Major Attacks 24  
       5.1.2  All data: A panel study with individual fixed effects  ............ 29  
   5.2  Exploring the Mechanism: Demotivation or Deterrence?  .......... 32
What are the consequences of committing violent attacks for terrorist organizations? As a result, do they attract more followers? Scholars and pundits alike generally refer to the existence of two opposing effects in their description of the consequences of violent terrorist activity: a mobilization effect and a discouraging effect (Kydd and Walter, 2006). Indeed, on the one hand, attacks may aim at solidifying and broadening the base of supporters within the terrorist organization. On the other hand, attacks are also a tool to show their capacity and infuse terror among the populace. While this effect targets the general population, it may also induce those who were already sympathizers to update their knowledge about the capacity of the group or to change their behavior with regards to the organization out of doubt or fear. While the theoretical arguments about these two strategies of terrorist groups abound, there is little empirical evidence for either of them.

An empirical challenge for researchers is that these two processes, mobilization and discouraging, are likely to happen simultaneously. However, they should occur with varying degrees of importance at different layers of the group-individual relationship. For instance, the chance that one of the two effects will affect their leaders is likely to be distinct from the influence an attack may have on an average member of the population. Consequently, we argue that the relevant question should not be whether a particular effect takes place at all, and not even if one effect dominates over the other, but which effect is dominant in each particular group of the population.

This paper begins to fill the empirical gap in the literature by providing an answer about which effect dominates among those segments of the population that are on the cusp of being outsiders and insiders regarding a terrorist organization. We do so by exploring the dynamics in those groups relative to the currently most famous terrorist organization, the Islamic State of Iraq and Syria (ISIS), on their most important communication and recruitment tool: Twitter.

In particular, we argue that the number of followers of ISIS-linked Twitter accounts reveals crucial information about the internal mobilization and demobilization dynamics
regarding the relationship between the terrorist organization and the society (Acosta, 2014; Bueno de Mesquita and Dickson, 2007; Kydd and Walter, 2006). For this, we first classify the followers into the following categories: 1) leaders; 2) active supporters (active members who unconditionally support the organization); 3) passive supporters (latent sympathizers who may either turn into active supporters in the future or withdraw their support from the organization); and 4) observers (non-supporters who follow the group to acquire information, such as media reporters, or individuals fighting against ISIS, but who would never support the organization). The former three groups are the insiders of the organization in the sense that they “follow” the organization and have shown a degree of sympathy towards it; the fourth group – as well as a fifth group composed of the entire population of Twitter non-followers – constitute the outsiders.

We argue that a terrorist attack is a shock that involves a re-positioning of some individuals across these five categories, some people moving upward in their commitment to the organization – what we call mobilization effects –, and some people moving downward in their expressed dedication to the organization – what we call discouraging effects. Further, they can be divided by whether these movements are internal, meaning, they occur among the insider group (e.g., already supporters becoming more or less supportive), or external, in other words, across the insider-outsider boundary (e.g., non-followers becoming passive supporters or passive supporters becoming non-followers). Importantly, the external processes of mobilization and discouragement determine the size of the Twitter audience; that is, the number of people that can be regularly reached by the organization for propaganda and recruitment purposes. Because of their strategic importance for terrorist organizations, this paper empirically focuses on the effects of terrorist attacks on the external kind of mobilization and discouraging effects.\(^1\)

Focusing on these external processes, we can see that overall changes in the number of followers before an attack and after an attack reflect three movements across categories.

\(^1\)Henceforth, we use the terms of discouragement, and mobilization to refer to their external components only.
First, non-followers may become observers. Second, non-followers may turn into insiders. Third, passive supporters may stop following ISIS-related accounts. While the first two movements imply an upward shift in the number of Twitter followers after a terrorist attack, the last movement implies a reduction in the number of followers. We empirically estimate whether the discouraging effect (movement of type 3) dominates the mobilization effect (movement of type 1) and attention effect (movement of type 2), by estimating the causal impact of ISIS’s terrorist attacks on the number of followers in ISIS-related Twitter accounts.

To test the theoretical inequality of interest empirically, we have implemented an automatic routine that collected the daily reports made by Anonymous on ISIS-related Twitter accounts and extracted key information on these accounts over 127 days – from March 14 to July 22, 2016. Altogether, we collected information on 300,842 observations, account-date data points, on 13,321 unique Twitter accounts. Then, we merged this dataset with data on real-world data on terrorist attacks around the world that occurred during our period of study. Using this constructed dataset, we examine whether terrorist attacks linked to ISIS exert an impact on the number of followers of ISIS-related accounts by combining observational and quasi-experimental research designs.

Different methodological perspectives provide a strong empirical regularity: ISIS’s terrorist attacks decrease the number of followers of ISIS-related Twitter accounts in the aftermath of an attack. The estimated effect regarding Twitter followers is both statistically and quantitatively significant. We employ two empirical strategies with identical conclusions. First, we exploit the panel structure of our dataset to examine the effect of the intensity of a terrorist attack on the number of followers in their accounts using several model specifications, including random and fixed effect models to account only for variation in the number of followers within each account. By doing this, we observe that terrorist attacks precede fluctuations over time in the number of followers in ISIS’s terrorist attacks; this is true especially for attacks that occurred on European soil as compared to other attacks in Asia and

\footnote{For our empirical analysis, we will only be able to use 239,434 account-date observations on 10,554 accounts to ensure the quality of our dataset. For additional details, see the data section of the paper.}
Africa for which the effect is still negative and significant, but smaller in magnitude. Second, we have also tested our main hypothesis by using an interrupted time series analysis. In this case, we explore the discontinuous changes in the number of followers around two major attacks in our dataset: the Brussels bombing on 22 March 2016 and the attack in Nice on 14 July 2016. Both analyses consistently point to the negative effect of the attacks on the number of followers of the terrorist organization on Twitter.

Our core finding provides evidence of a discouraging effect of terrorist attacks on the followers who support the organization, the third class of movement in the typology specified above. This effect is particularly strong considering that we only observe the net effect: the discouragement effect is discounted for the opposite effect of both the mobilization and the attentional effects among previous non-followers who become observers. Yet, the discouragement effect is not homogeneous, but it is constituted of two components: a demotivating effect, the action of un-following the organization due to a change in supporters’ political beliefs in the aftermath of an attack; and, a deterrence effect, the act of un-following the organization to avoid being tracked after an attack. To evaluate the distinct mechanisms, we examine whether the discouraging effects are, as the deterrence mechanism would suggest, stronger in accounts located in countries where governments’ counter-terrorist activities are more active, that is, Western countries and/or the region of the attack. Instead, we find that the effect of terrorist attacks has no effect on the number of followers in the sub-sample of accounts that we can identify to be in Europe. Hence, we conclude that the demotivating effect, a change in beliefs, is the most plausible explanation to the negative effects of terrorism on the organizations’ supporters.

Overall, this paper speaks to several strands of literature. First, we first offer a critical theoretical account on the types of effects that terrorist attacks may have on terrorist organization, as well as its relationship with the population. Second, we present, to the best of our knowledge, the first systematic study on the impact of terrorist attacks on the (de-)mobilization of terrorist supporters. Though there is some prior empirical research on
the consequences of terrorism on some relevant political outcomes, including people’s attitudes and ideology (e.g., Davis and Silver, 2004; Gadarian, 2010; Getmansky and Zeitzoff, 2014; Peffley, Hutchison and Shamir, 2015) and governmental policies (e.g., Abrahms, 2012; Thomas, 2014), the study on its impact on the organization and its potential supporters has thus far not been driven by empirical data. Secondly, our work contributes to the growing literature that employs big data to answer long-standing questions in political science (e.g., Barberá, 2015; Barberá et al., 2015; King, Pan and Roberts, 2014a,b, 2016) by using a novel dataset on ISIS-related Twitter accounts. In addition, our work connects to the heated debate in political science on whether big data and causal inference are contradictory trends by employing big data in a quasi-experimental framework (Monroe et al., 2015).

The rest of the paper is organized as follows. First, we provide the literature review on the effects of terrorist attacks on the population. Then, we present a modeling strategy for the net effect of the mobilization and discouraging effect among those who are susceptible of crossing “the borderline” between insiders and outsiders. Next, we turn to our research design where we describe our data extraction method, the nature of the two datasets that we combine, and our empirical strategy. After that, we show our results from the panel data, the interrupted time series analysis for two major attacks illustrating our estimated effect, as well as evidence for our suggested mechanism. Finally, we conclude the paper.

1 The Impact of Terrorist Attacks on Twitter Followers

In this section, we first discuss some of the theoretical approaches that have been used to understand terrorist activity, with an emphasis on theories about how ISIS seeks to influence their base of supporters and the general population through terrorist attacks. Second, we discuss the use of Twitter in ISIS’s recruitment and propaganda strategies, as well as a plausible conceptualization of the types of users in Twitter with regards to its relationship
with ISIS. Both sections serve as the building blocks for our modeling strategy that connects our theory to our empirical Twitter data.

1.1 Terrorist Attacks as Strategies of Mobilization and Discouragement

There are several possible explanations for the relationship between terrorist attacks and the size of an organization’s audience. While we acknowledge that variation in the terrorist activities of the political groups may be explained by a number of factors, including the psychology of individual terrorists (e.g., Horgan, 2005; Lankford, 2010; Post et al., 2009; Victoroff and Kruglanski, 2009), adherence to religious ideals (Pargament, Magyar-Russell and Murray-Swank, 2005), socialization processes (Turk, 2004), contextual socioeconomic factors (Krueger and Malečková, 2003; Mitra et al., 2008; Piazza, 2011), or their organizational structure (Asal and Rethemeyer, 2008), we construct our theoretical prediction on the basis of the scholarly body of literature that views terrorist attacks as a rational choice to achieve political aims.

Doubtlessly, the most influential view on the purpose of terrorist attacks in political science suggests that terrorist groups use violence as a costly signal to show strength and capacity. The scholarly debate focuses on who is most-often targeted by the signal and its effectiveness. Since the ultimate objective of violence is the extraction of policy concessions, governments and people both at home and abroad are most often the targets of violence. While the effectiveness of groups’ strategies is a matter of scholarly debate,\(^3\) Kydd and Walter (2006) suggest that groups pursue five goals via the costly signaling of terrorist attacks: 1) attrition: attacks may aim to persuade governments to end a conflict by increasing

\(^3\)Using all terrorist attacks between 1980 and 2003, Pape (2005) finds that half of terrorist attacks have achieved some policy concessions with the targeted states. However, Abrahms (2012) finds evidence contrary to this claim. Most recently, Thomas (2014) uses extensive empirical evidence on terrorist activities and war outcomes, and finds support for the thesis that terrorism is rewarded with both greater political concessions and bargaining power throughout the peace process. Besides policy concessions, terrorism may also have (un-)intended institutional and economic consequences (e.g., Abadie and Gardeazabal, 2003; Indridason, 2008).
the costs of continued fighting; 2) intimidation: violence seeks to convince the people to change their behavior because the costs imposed on them are too high (e.g., begin punishing the government); 3) provocation: terrorists aim to trigger a violent policy response that encourages even greater mobilization in support of the group; 4) spoiling: the aim of the attacks is to disrupt a peace process; and 5) outbidding: the attacks aim to increase the people’s support for that group in the presence of competitor groups. Though each goal has distinctive characteristics, all of them share a similar underlying mechanism: terrorist attacks change the political position of the organization through the provision of new information about the organization to either the government, the people or both.\(^4\)

Regardless of the goal they pursue, terrorist attacks may have two effects on people that witness violent acts: 1) infuse terror in the population, and/or, 2) consolidate their base of supporters.\(^5\) On the one hand, terrorist violence is seen as a strategic mobilization tool that aims at ensuring a “self-sustaining rate of [political mobilization]” (Acosta, 2014), solidifying the loyalty of the already militant members, and broadening the base of supporters to ultimately institutionalize the organization (Acosta, 2010, 2014, 2016; Bueno de Mesquita and Dickson, 2007). Not only does the nature of the attacks signal strength to the people to achieve their ends, but it also conveys relevant information to their supporters about the kind of strategies used by the organization, the character of their members, and the goals of the organization (Bueno de Mesquita and Dickson, 2007; Hoffman and McCormick, 2004). The mobilization strategy can be divided into two types: internal and external mobilization. \textit{Internal mobilization} implies an increase in the commitment and loyalty of those who are already members of the organization. \textit{External mobilization} refers to a process of broadening the base of supporters by bringing new members – either active or passive – into the organization.

\(^4\)Notice that even if some terrorist groups may not have full control over all violent activities on the ground, their execution on behalf of the organization – regardless of the organization’s prior consent – should be sufficient to imply their political and social consequences.

\(^5\)These two mechanisms are simply a summary of the shared mechanisms within each of the five strategies mentioned above.
On the other hand, the new information about the organizational capabilities, the character, and the goals of the organization also serve the purpose of infusing terror among the general populace with effects at home and abroad. When organizations have territorial control, such as ISIS in Iraq and Syria, or Boko Haram in Chad, Niger, and northern Cameroon, organizations may find terror as a useful device to signal internally their organizational strength and their capacity to maintain control over law and order within their territory. Among other consequences, this may allow them to appease dissident activity. In addition to the domestic effects of violence, frightening the international population may also be an alternative goals for anti-system terrorist groups such as the Islamic State. First, it may seek a change in the position of societies toward their government and, thus, the response of the government with a policy concession targeted to the organization, such as the withdrawal of military forces from the areas of influence under control of the organization (Indridason, 2008; Merolla and Zechmeister, 2009a,b). And, second, frightening the foreign population may also serve as a form to trigger pressure from the citizenry to their government and, thus, provoke governments to engage in indiscriminate military attacks against civilians, which may then allow them to gain supporters among their domestic citizenry (Bueno de Mesquita and Dickson, 2007; Carter, 2016). However, terror is not an effect that can be targeted exclusively to some segments of the population, but it affects everyone. Consequently, violent activities also convey information to mild supporters about the characteristics of the organization, which may lead to two distinct effects: greater support (internal mobilization effect) or withdrawal of their expressed support (external discouraging effect).

1.2 The Islamic State on Twitter

Consistent with both of these theories, an increasing number of terrorist groups make an intensive use of mass media and social media to disseminate their messages. The availability of these new technologies has made it far easier to distribute the groups' films and images: terrorist attacks and violent scenes are reproduced live uncensored, which has led to an
unprecedented stream of online violence. Besides the use of regular television channels to promote their films in the news, terrorist organizations have begun to be highly involved in a number of decentralized digital platforms through social media networks like Twitter or Facebook; peer-to-peer messaging apps like Telegram and Surespot; and, content sharing systems like JustPaste.it (Clark, 2016). While this has been the case for a large number of terrorist organizations worldwide, the Islamic State has intensively used Twitter for both propaganda and recruitment purposes. In this regard, FBI Director James Comey argued in July 2015 that ISIS’s strategy consists in “broadcasting on Twitter, get people to follow them, then move them to Twitter Direct Messaging”. After a short screening of whether the follower is a likely recruit, then “they’ll move them to an encrypted mobile-messaging app so they go dark to us” (Aspen, 2015).

From this perspective, we could classify Twitter users into five types according to their relationship with ISIS: 1) ISIS’s leader members: they constitute a number of owners of ISIS-linked accounts and form a share of the followers of the ISIS-related accounts; 2) ISIS’s active supporters (active engagement): they may or may not have ISIS-related accounts themselves, but they are surely a share of the followers of ISIS-related accounts; 3) ISIS’s passive and likely-supporters (passive engagement): they do not have ISIS-related accounts, but they are passive consumers of ISIS’s information, are susceptible to converting into active supporters, and constitute a great share of the friend-followers of the ISIS-related accounts; 4) the observers: they are the people who do not support, but follow their accounts because they are consumers of ISIS’s information (e.g., media reporters who follow the latest news in Syria through ISIS’s Twitter accounts) or people fighting ISIS; and, 5) the non-followers: they are all other users on Twitter who do not follow and do not support ISIS’s activities currently, yet all of them could turn into ISIS’s passive or active supporters in future periods. This group constitutes the vast majority of Twitter users.

With this categorization of Twitter users, we conceptualize an attack as a shock that reshuffles the number of people in each of the five types of groups. There are two oppos-
ing forces that define movements across categories: the mobilization effect, which leads to
movements toward the group of the leaders (e.g., non-followers to passive, passive to active,
active to leader, etc.), and the discouraging effect, which leads to movements toward the
non-followers (e.g., leader to simply active, active to passive, passive to non-followers, etc.).

For reasons of empirical identification, we further classify the types of mobilization
and discouragement as internal and external. Table 1 reports the dynamics that occur
before and after a terrorist attack at the individual level both internally and externally,
and Table 2 reports the external processes alone at the level of aggregate effects, those
which will be observable to us. Internal effects refer to changes across categories within the
organization, which can be either positive – stronger commitment – or negative – weaker
commitment. If the change is a stronger commitment within the organization, then this is an
internal mobilization effect. This occurs when an individual that was passive becomes more
active and, for instance, becomes more engaged in terrorist activities, such as propaganda
diffusion. It would also be an example of internal mobilization when an active member
increases his or her loyalty by taking positions of responsibility within the organization.
However, movements can also be defined as internal discouraging effects when people decide
to reduce their commitment. For instance, ISIS fighters returning home may move from
being a leader to a mere active or passive supporter. Similarly, a member who used to be
highly engaged may become less active for a number of reasons.

[Table 1 about here]

[Figure 2 about here]

External effects refer to movements to and from following ISIS-related Twitter accounts.
External encouragement effects have two components. The most relevant of the two for the-
tories of terrorism is the mobilization effect. This occurs when a non-follower becomes an
explicit supporter of the organization by following it on some of their related accounts. Nat-
urally, we should expect that the external mobilization process is more likely to occur across
adjacent categories because the process of radicalization is not immediate. Therefore, we be-
lieve that an external mobilization process is likely to take place when non-followers become passive supporters of the organization. In a second place, attentional effects define aggregate movements of observers toward ISIS-related Twitter accounts; they are non-followers who begin following their accounts with the simple purpose of obtaining first-hand information on the organization and their activities (e.g., media reporters), yet they have no chance of becoming supporters of the organization.

There may also be an external discouragement process, in which passive supporters update their beliefs over the organization and, some of them may realize that the organization does not fulfill their desires, so they withdraw their mild support (a demotivating effect). In addition, some previous ISIS’s followers may become discouraged from following ISIS-related accounts out of fear of being tracked by intelligence agencies (a deterrence effect). Both processes are reflected in the movement to the category of non-followers. In short, our first hypothesis reflects the existence of these two rival sub-hypothesis and, thus, the existence of an encouragement or a discouragement effect.

The nature of the treatment effect should not be regarded as the terrorist attack alone, but we should conceptualize in a broad sense. To the discrete event of the attack, the actual treatment effect includes all of the causal chain that a terrorist attack triggers, which includes, for instance, the reactions to the attack by other citizens, media, politicians, institutional figures, and the like. This compounded treatment is obviously larger as the number of casualties increases and the geographic location of the attack is in a Western country. Hence, if the attack is the actual cause a change in the ISIS-related number of followers on Twitter, then we should also expect that these attacks exert larger effects, whether positive or negative, when they occur in Western countries compared to elsewhere. This is our second hypothesis.

In addition to establishing the net effect of terrorist attacks on the number of followers, we tease out the logic behind demotivation and deterrence to explore the mechanism of the negative effect. The crucial distinction between demotivation and deterrence is whether the
change in individuals’ behavior is the result of learning from the attacks and, thus, a shift in beliefs; or, the consequence of fear or threat of prosecution. Though we do not have data to directly observe neither of these two mechanisms, we lay out a key empirical observation for the deterrence effect, and then subject it to examination.

According to the deterrence effect, terrorist attacks increase the beliefs among followers that their online behavior may become noticeable and tracked by security agencies. Thus, ISIS’s supporters, whether passive or active, may become dark on Twitter by un-following or shutting their accounts after a terrorist attack, which would explain the decrease in the average number of followers that we observe. If the deterrence effect dominates our negative result, then we should see that those followers located in the region of the attack and/or in countries with stronger counter-terrorist responses, namely, Europe and the United States, should be more likely to drop from following ISIS-related accounts in the aftermath of an attack terrorist attacks, compared to the rest of our sample. Testing this hypothesis means testing for the relevance of the demotivation mechanism conditional on finding a negative effect in our main hypothesis.

In sum, the logic of the external processes gives rise to the following expectations:

**Hypothesis 1 (main hypothesis: encouragement versus discouragement)**

- **Hypothesis 1.1** The number of followers on ISIS-related accounts increases (decreases) after a terrorist attack.

- **Hypothesis 1.2** The effect of terrorist attacks in either Europe or the United States is stronger than the effect of attacks occurring elsewhere.

**Hypothesis 2 (conditional on discouragement: demotivation mechanism)**

The number of followers on ISIS-related accounts decreases after a terrorist attack, yet the effect is not stronger among European-only accounts compared to accounts located.

These external processes are crucial for information spread and recruitment purposes because they determine the size of the Twitter audience. Twitter audience can be defined as
the set of people who can be reached through the network of ISIS-related accounts. Hence, they are more likely to be susceptible to ISIS’s propaganda and to be recruited than the general population. In terms of our classification, the Twitter audience is composed of leaders, active and passive supporters, and observers. Therefore, the net effect of the two opposing external processes, mobilization and discouragement, determines changes in the size of the Twitter audience.

2 Formal modelling of the response to an ISIS attack

The purpose of this section is to analytically identify the sources of the change in the number of the followers of the reported accounts as a response to the ISIS attacks. As already proposed, the observed number of the followers consists of the group of leaders, as well as actively and passively involved people who are not in a leadership position; furthermore, some followers are just “observers”: terrorism fighters, analysts or journalists. Hence, we need to model the mechanism of transition between those categories and non-followers. In this section, we provide a formal justification for our main hypothesis and show that finding a negative effect of the ISIS attack on the number of Twitter followers indicates that the external discouragement effect of the attacks is stronger than the external mobilization effect.

Importantly, the following model contains five categories while empirically we can discriminate only between followers and non-followers. Having a richer classification in the formal model than in the data is crucial as it enables to underline three key points. First of all, this model proposes a simple way to look into the general dynamics of the online extremist mobilization on Twitter. Second, as a result of the first point and crucially for our main argument, it allows showing explicitly that the existence of the hidden internal mobilization dynamics does not contradict our hypotheses about the external mobilization dynamics given our data. Third, the model explicitly links the broader formal setting to our causal empirical claims, providing additional evidence for them.
2.1 Individual transition model

Based on the classification in the previous section, here we describe the individual likelihoods of the transition between the categories. The stochastic matrix (Table 2) formalizes our substantive assumptions about the types of the followers and their likely responses to a terrorist attack by ISIS. We design the movements across the categories as a Markov process with no memory - only the current state affects the possible dynamics. Clearly, from the substantive perspective, the history influences the current position as well. However, since all that information is embedded in the present state the lagged variables are omitted. The model has five individual states, two of which are fixed-points: “leader” and “observer.” Meanwhile, three other states enable movements up or down one category regarding the individual mobilization.

[Table 2 about here]

Each row shows the non-zero probabilities of the member of a particular group to transition to another category or to stay in the same category: \( \alpha_0 + \alpha_1 + \alpha_2 = 1, \beta_0 + \beta_1 + \beta_2 = 1, \) and \( \gamma_0 + \gamma_1 + \gamma_2 = 1. \) First, we assume that the current ISIS’s leaders do not respond in any way to the attack. That is why the only non-zero value on the first row is on the diagonal. Contrary to them, the active supporters might become more or less mobilized as a response to the attack, or they might keep the same level of interest. Most importantly, we assume that they will be still among the detected followers even if their support becomes milder. The categories providing the variation in the observed numbers of the followers are passive supporters and former non-followers. Talking of the passive supporters, some of them might radicalize. Meanwhile, some may dislike what they see or become afraid of being tracked as connected to ISIS and stop following ISIS. The observers, who are most likely ISIS enemy fighters or journalists following ISIS, will not change their behavior as a response to the attack. Finally, some of the non-followers might become passive supporters as the terrorist attack makes ISIS more visible; this would imply a broadening of the base of supporters of
the organization – what we call the external mobilization effect.

2.2 Group comparative statistics

Let’s denote: \( x = (x_1, x_2, x_3, x_4, x_5) = \) (leader group, active, passive, non-followers, observers). Our data enables to observe the total number of the followers of the reported accounts. This is the exact upper bound of the actual total number of the people in categories 1-3 and 5, since some reported Twitter accounts may have common users following them. We assume that the correlation between the observed upper bound and the total number of the distinct followers is approximately the same before and after the attack. Hence, if we observe the decrease in the exact upper bound after the attack, it indicates a decline in the total number in categories 1-3 and 5.

To sum up, the total observed number of the followers as: \( x = x_1 + x_2 + x_3 + x_5 \), where \( x_i \) denotes the number of the people in group \( i \) (\( x_4 \) are the non-followers before a terrorist attack). We need to estimate \( E(x^*) \) the expected number of the observed followers after the attack based on Table 2. Given that every follower is independent of one another, the estimates for the categories after the attack are:

\[
E(x^*) = A'x' = \begin{pmatrix}
  x_1 + \alpha_0x_2 \\
  \alpha_1x_2 + \beta_0x_3 \\
  \alpha_2x_2 + \beta_1x_3 + \gamma_0x_4 \\
  \beta_2x_3 + \gamma_1x_4 \\
  \gamma_2x_4 + x_5
\end{pmatrix}
\]

Summing up, what we observe after an attack:

\[
x^*_1 + x^*_2 + x^*_3 + x^*_5 = x_1 + x_2 + (\beta_0 + \beta_1)x_3 + (\gamma_0 + \gamma_2)x_4 + x_5
\]

Hence, the observed change is:
\[ \sum_{i=1,2,3,5} x^i - \sum_{i=1,2,3,5} x = \delta = -\beta_2 x_3 + (\gamma_0 + \gamma_2) x_4 \]  

(1)

where \( \beta_2 x_3 \) is the \textit{external discouraging effect}, \( \gamma_0 x_4 \) is the \textit{external mobilization effect}.

Hence:

\[ \delta < 0 \implies \{ \gamma_2 x_4 > 0 \} \implies \beta_2 x_3 > \gamma_0 x_4 \]  

(2)

Importantly, because of \( \gamma_2 x_4 \neq 0 \) if \( \delta > 0 \) we are not able to evaluate the relation between the \textit{external discouraging effect}, \( \beta_2 x_3 \), and the \textit{external mobilization effect}, \( \gamma_0 x_4 \).

Three possible empirical outcomes conclude this formal section. If we observe an increase in the total number of followers of the reported accounts, it implies that the combination of the mobilization and the attention effect dominates the discouraging effect \( (\gamma_0 x_4 + \gamma_2 x_4 > \beta_2 x_3) \). If the number of followers is not altered, then it must be the case that the discouraging effect cancels the other two effects \( (\gamma_0 x_4 + \gamma_2 x_4 = \beta_2 x_3) \) and we still could claim that \( \beta_2 x_3 > \gamma_0 x_4 \). Finally, if we observe a decrease in \( \delta \), the discouraging effect of the attacks dominates their mobilization effect \( (\gamma_0 x_0 < \beta_2 x_3) \).

3 Data

3.1 Information on ISIS-linked Accounts

The data used in this paper were obtained by following “the reports” of the anti-ISIS Twitter bots set up by the international hacker initiative \textit{Anonymous}. This hacker group declared cyber war on ISIS almost immediately after the ISIS attack in Paris on November 13, 2015. The horrible event took place on six sites of this European capital and eventually took the lives of 130 people, leaving 368 injured (Mosbergen, 2015; Holmes, 2015; Ellyat, 2015). Having said to ISIS “We will let you down” in their famous video,\(^6\) the initiative started to work in multiple directions, one of which was to report ISIS-related accounts in

\(^6\)For the full video, see https://www.youtube.com/watch?v=5UjIqw9fyk.
Twitter.

This paper makes use of 300,843 reports from the anti-ISIS *Anonymous* bots collected over the period from 03/16/2016 to 07/20/2016. Each report is a Twitter account. Most of the accounts in this data set have been reported multiple times and eventually got suspended (hence, there is a clear life-cycle of the accounts in the sample: they were reported the first time, then they were reported a couple of more times, then they were suspended). Each time when a certain Twitter account was reported, its profile information was collected. The gathered data is an unbalanced panel, in which \( n \) is a reported Twitter account and \( t \) is the date of this report. Overall, the final dataset has information on 13,321 unique accounts.\(^7\)

Crucially for our analysis, each entity in the data contains information on the number of followers, friends, the date of the report, and the language of the account.\(^8\) Our dependent variable is the number of followers in an account in date \( t \). To test our third hypothesis, we need to create a dummy variable that takes into account whether an account is located within Europe or North America, or elsewhere. While we cannot geo-locate the accounts with their profile information, the language of the accounts can be used to identify those accounts that have a high probability of being located in Europe if their language is exclusively spoken within Europe and North America.\(^9\) Thus, we divide all our accounts in two types: 1) accounts whose language is only spoken in countries within Europe and North America

---

\(^7\)Our sample for analysis excludes observations that have extreme activity rates. Accounts with a high ratio of following accounts over followed accounts are likely to be bots, and accounts with a low ratio of following accounts over followed accounts are likely to be media. Because they may behave differently than regular accounts and their behavior may be too influential for our results, we have chosen to exclude the lowest and the highest decile. Our final sample size is, thus, 239,434 user-date from 10,554 accounts. While trimming the dataset on the extremes aims at cleaning the data of extreme observations, the findings presented in this paper are the same if we include all of the observations.

\(^8\)The raw data collection process was performed using a couple of specific Python scripts running on a remote server and gathering information 24/7. Then, the data refinement and aggregation was performed on a stand-alone machine via a C# .Net application in Visual Studio 2015; the application was developed particularly for this project.

\(^9\)The language is automatically generated by Twitter when the user sets up an account.
alone,\textsuperscript{10} and, 2) all other accounts.\textsuperscript{11} This information allows us to construct an *European-only language* dummy that we will use to test our third hypothesis.

\subsection*{3.2 The Cumulative Lagged Measure of Terrorist Attacks}

We complement the information extracted from the accounts reported by *Anonymous* with data on daily terrorist incidents linked to ISIS. An incident is included in our dataset if it has been connected to or has been said by reliable sources to be inspired by the Islamic State of Iraq and Syria (ISIS), also known as Islamic State of Iraq and the Levant (ISIL) or Daesh. Our online appendix includes the news sources and further details on each of the incidents considered in our dataset.

Our assumption is that the effectiveness of either a discouraging or a mobilization strategy should be greater as the impact of the attack is larger. Thus, this impact factor is likely to be related to two components: the number of casualties and the location of the incident. For this reason, we construct two independent variables for each day in our period taking into account these two parameters. Firstly, we compute the number of deaths from terrorist attacks linked to ISIS that occurred in the United States and Europe. Secondly, we do the same for the number of deaths outside the United States and Europe. Hence, our independent variables do not reflect just whether there is an attack on a particular day, but they weight each attack for the number of deaths that it generates, and differentiate between those in Europe and the US and elsewhere. If our empirical strategy reflects our theoretical reasoning, then we should observe that the effect of the number of victims in the US and Europe should be larger than the effect of casualties located elsewhere geographically.\textsuperscript{12}

\textsuperscript{10}Czech (< 0.01%), Danish (0.16%), Dutch (1%), Finnish (0.04%), German (1%), Italian (< 0.17%), Norwegian (0.04%), Poland (< 0.01%), Russian (0.71%), Swedish (0.25%), and Turkish (< 4.40%); total accounts with an European-only language constitute a 7.2% of all accounts. Also, we should note that recoding Russian and Turkish user accounts to the category of non-European-only languages does not change any of the results presented in this section.

\textsuperscript{11}The majority of users whose language is not exclusively spoken in Europe and North America fall in one of the following categories: Arabic (45.7%), English (32.5%), French (10.7%), Indonesian (2.5%), and Spanish (0.47%).

\textsuperscript{12}There are no attacks in the US in our dataset, although we include a mention to the US throughout the
However, the assumption of an instant effect of a terrorist attack lasting only one day is unrealistic. Hence, a smoothing temporal approach for effect is a logical decision. Indeed, media coverage and news stories linger for some days after an attack. Many people may find out about events one or two days later. Consequently, we should expect to have the consequences of an attack on the same day of the attack and over its immediate aftermath, although, probably, decreasing in magnitude. For this reason, we have chosen to implement a smoothing decreasing curve after the shock by applying a discount factor of 50%, 75%, and 100% to the number of deaths. For each day in our dataset, we compute the cumulative mortality parameter, \( cd_t \); it aggregates all previous deaths from the terrorist attacks with a discount factor:

\[
\begin{align*}
    cd_t = \sum_{i=1}^{I} \frac{\text{deaths}(i)I(t \geq t(i))}{(1+r)^{t-t(i)}},
\end{align*}
\]

where \( i \) is a terrorist attack on day \( t(i) \) and \( \text{deaths}(i) \) is the number of deaths it caused; \( r \) is the discount rate applied, and \( (t - t(i)) \) is the number of days since the incident. This is summed for all incidents before \( t \).

For example, we estimate the cumulative mortality parameter around the Brussels bombings on March 22, 2016, with \( r=0.5 \) as having values of 0, 36, 24, 16, 10.67, and so on, for the days between 21th to 25th March, respectively. Figure 1 shows how this discount pattern parallels the changes in the keywords “Brussels bombing” reported by Google Trends. On the date of the occurrence, Google trends reports a value of 100 (its standardized base value), a decrease to 7 one week later, and a further decrease to 0 two weeks later, for both attacks. Similarly, if we applied a discount factor of, for example, 50% to an event with 100 deaths – as the base in Google trends for comparison –, then the trend would be 100 in paper because we would have included attacks in the US in a separate category with Europe. In addition, we refer to attacks in European soil to those attacks within our period that occurred in Belgium, France, Germany, and Turkey. However, if we exclude Turkey from this category, the results for the variable of attacks in Europe become even stronger.
the same day, 5.85 a week later, and 0.34 two weeks later.\footnote{The correlation of Google attention to the event and our cumulative value is above 0.90 within a one-month window around the event. Other attacks in the sample do not differ in the evolution of their attention over time, but some of them had obviously less overall attention. It is worth noting that choosing a greater discount parameter does not alter any of the results presented in the paper.}

[Figure 1 about here]

To illustrate the incidence of violence linked to ISIS within our period of study, Figure 3 reports the cumulative lagged measure of deaths from terrorist attacks linked to ISIS from our initial extraction date (03/16/2016) to the end of the data collection period (07/20/2016). Altogether, there are two major attacks in European soil: on 22nd March in Brussels, Belgium; and on 14th July in Nice, France.\footnote{The attack in the Attaturk airport in Istambul, Turkey, is another attack that is coded as in European soil, although it is unclear whether it should be computed in Europe or outside Europe. Notice, however, that including it in the dataset as Europe or outside Europe does not alter any of the results presented in the paper.} Outside European soil: the major attack was the two coordinated bomb attacks in the Karrada district in Baghdad, Iraq, on the 3rd of July, with over 300 deaths.

[Figure 3 about here]

4 Empirical Strategy

Exploiting the temporal variation in terrorist attacks throughout our period of study, we use a pre-post approach to estimate their effects on the number of followers in the ISIS-related Twitter accounts. More specifically, for account $a$ on day $t$, we estimate the models of the form:

$$Y_{at} = \beta_0 + \beta_1 cd_t + \alpha_a + \epsilon_{at},$$

where $Y_{at}$ denotes the log of followers of account $a$ on day $t$. As noted previously, $cd_t$ is the continuous variable that equals the number of terrorist victims in day $t$, smoothed with a discount rate of 50%, 75%, or 100%. The parameter of interest in our model equation is $\beta_1$.

Our main specification includes user-account fixed effects ($\alpha_a$). With regards to the plausibility of the identification condition, account fixed-effects are important because there
are strong differences across accounts – language, geographic location of the user, demographic characteristics, as well as other observed and unobserved covariates. The major danger of an empirical strategy without account fixed effects is selection bias in the sample before and after an attack. This can arise as a result of shifts in either the type of accounts that *Anonymous* targets, which may be different in a period before and after an attack, or shifts in the behavior of users, including the opening or closing of accounts. The inclusion of account fixed effects allow us to remove all time-invariant heterogeneity across accounts (e.g., average number of followers throughout the period, language, geographic location) and focus on the average variation within accounts over time.

For the interpretation of the parameter $\beta_1$ as a causal estimate of the effect of terrorist attacks on the number of followers, we need to assume that timing of the terrorist attacks is exogenous to the pre-attack number of followers in Twitter accounts linked to ISIS. While a violation of the assumption is more likely when using long periods of time, we believe the assumption of the exogenous timing of terrorist attacks is highly plausible, because our sample period is comprised of a short time span of only 127 days. For instance, the likelihood of a terrorist attacks in 2012 may be very different from that of a terrorist attack in 2016 for reasons that we do not fully understand. However, the chance of a terrorist attack in a given day within the 127-day period of our analysis – from 16th March to 20th July, 2016 – is likely to be approximately constant within short periods from the perspective of Twitter followers and non-followers. In practical terms, this means that the exact day of the attack is generally not known in advance.

In short, for the causal parameter that we estimate to be biased, not only does it require the chance of an attack to vary over time, but to vary as a function of the number of followers in the periods immediately before an attack, or without an attack.\textsuperscript{15} Thus, the parameter $\beta_1$ can be interpreted as a causal estimate of the effect of terrorist attacks on the number of followers of ISIS-related accounts because the uncertainty related to the precise timing and

\textsuperscript{15} When we refer to an attack day, we are actually referring to the attack day and a number of days in the aftermath of the attack that are affected by the attack
nature of the attack provides an identification mechanism for our empirical tests.

5 Results

5.1 Mobilization or Discouragement?

5.1.1 Case studies: An Interrupted Time Series Analysis of Two Major Attacks

The empirical section starts with the evaluation of the reduction in the number of followers after a terrorist attack by looking at two particularly important terrorist attacks that fall within our time period of study: the Brussels bombing on March 22 and the Nice attack on July 14th, both in 2016.

To assess the causality of the relationship, we employ an interrupted time series analysis, a type of regression discontinuity design (RDD) in which the running variable is defined by time (Morgan and Winship, 2014; Percoco, 2014; Shadish, Cook and Campbell, 2002). The terrorist data and the information on Twitter accounts is ideal for this approach because of the large number of accounts that we have collected in every single day throughout our period of analysis, which statistically empowers small bandwidths, and the well-defined moment of the attack.

In brief, there is a potential outcome \( Y_j \) for each account-observation \( j \) – the number of followers of an account; a treatment assignment variable \( W_j \) – observation observed in the pre- or post-treatment period –, which determines the potential outcome we eventually observe, so \( Y_j(W_j) \), i.e., either \( Y_j(1) \) or \( Y_j(0) \); a forcing variable \( t_j \), which is a running covariate – time to the attack in our case –; and, finally, a cutoff value \( c \), which establishes the interruption in the time series, which is set at 0 (the attack date in our case). Hence, the treatment assignment is given by:
\[ W_j = \begin{cases} W_j = 1 & \text{if } t_j \geq c \\ W_j = 0 & \text{if } t_j < c \end{cases} \]

Therefore, the primary quantity of interest is the immediate change in the number of followers represented by \( \tau \) in the following equation:

\[
followers_j = \alpha + \tau W_j + f(t_j) + \epsilon_j
\]

where \( followers_j \) indicates the number of followers in an account-observation, \( \alpha \) is an intercept of the model, \( Attack_j \) is an indicator for whether an observation falls in the the day of the attack or later, \( f(t_j) \) is a smooth function of the running variable time in days from the day of the attack, which can be positive or negative, and \( \epsilon_j \) is an error term. To estimate \( f \), we use a non-parametric estimation (local linear regression).

A crucial step in this approach is the decision about the bandwidth – in this application, the window period of analysis. In general terms, there is usually a trade-off between precision and bias. On the one hand, a strategy that uses a large bandwidth would allow us to precisely estimate a causal effect because it would use a large number of observations ranging from many days before and many days after the attack. The risk of this approach is the lack of validity due to omitted variable bias. Because it allows data points that are far from the attack to be used in the estimation of the outcome in both the control and treated period, differences across units may arise because they were generated in different environments. On the other hand, we would like to minimize the bias of our estimate by subsetting our temporal focus to a narrow temporal window, and thereby removing those observations that are far from the day of the attack. Yet, the general downside of this strategy is that we end up working with a reduced sample size, which reduces the precision of our estimates.

In our model specifications, we choose a baseline bandwidth following the parameters suggested by Imbens and Kalyanaraman (2012) (IK) and test the sensitivity of the estimates by using different bandwidth, ranging from the bandwidth suggested by IK to only one day before and after the attack. In the estimation procedures, we use triangular kernel, as
recommended by (Fan and Gijbels, 1996; Lee and Lemieuxa, 2010) and (Lee and Lemieuxa, 2010), which is equivalent to estimating a weighted linear regression over the interval on both sides of the cutoff point, giving more weight to those observations closer to the cutoff point.

### Attack on Brussels, Belgium

On 22 March 2016, two blasts hit the main terminal of Zaventem international airport, in the north-east of central Brussels, and another explosion struck the Maelbeek metro station in the area where several European institutions are located. These three coordinated suicide bombings killed thirty-two civilians and three of the perpetrators, and more than 300 people were injured. Shortly after the explosions, ISIS claimed responsibility for the attacks in a statement released via Twitter, Telegram, and other social media.

To evaluate the impact of the attacks on the number of followers, we subsetted our dataset to only those observations that are observed more than once and fall within the window of 15 days before and after the attack.\(^\text{16}\) As our data collection procedures were initiated on March 16, the pre-attack period is six days before the attack and the post-attack period is 15 days after the attack.\(^\text{17}\)

Table 4, Panel A, presents the main estimates of the impact of the terrorist attack in Brussels on the number of followers in the accounts linked to ISIS. Table 4 shows the findings from our above-mentioned estimation equation by local linear regression, using the number of followers as the dependent variable. As indicated above, our treatment variable is \(\text{Attack}_{i}\), a dummy that indicates whether the observation is in a pre-treatment or post-treatment period. Therefore, the coefficient \(\tau\) on \(\text{Attack}\) captures the effect on the number of followers of being in the pre-attack period versus the post-attack period, i.e., it measures the effect of crossing the threshold – the attack date, from left to right – on the number of followers. Column (3) provides the estimated effect based on the preferred specification, which uses

---

\(^{16}\)We excluded all observations with the extreme number of followers – the top 1% percentile. However, the results are not driven by these decisions.

\(^{17}\)This decision of choosing a specific time window is not relevant analytically because the bandwidth used is systematically narrower than this time range.
the bandwidth suggested by Imbens and Kalyanaraman (2012).

Under this specification, the attack in Brussels decreased the number of followers by 37 followers, which is equivalent to a 9.9% decrease from the mean number of followers of the twelve-day period – six days to the left and six days to the right on the attack’s timeline. Importantly, the estimated negative effect is consistent significantly and quantitatively similar across different bandwidth specifications, whether doubling or halving the IK bandwidth. Moreover, even if we focus on the two-day period centered around the attack day (bandwidth = 1), one day before and the attack day, the effect is also significantly negative. Figure 5a graphically shows the discontinuous jump around the attack date. This provides substantial evidence of the causal impact of the Brussels attack on the number of Twitter users following ISIS-related accounts.

This subsection explores whether we see a similar effect for the largest terrorist attack on Western soil during our period of study. On the night of July 14, 2016, a man driving a 19-ton refrigerated truck and carrying an automatic pistol deliberately drove into crowds that were celebrating the Bastille Day on the Promenade des Anglais in Nice, France. The attack resulted in the deaths of 86 people and 434 injuries. On the morning of July 15, 2016, a few hours after the attack, the French president, Francois Hollande, addressed the nation in a televised speech, stating that the attack was terrorist in nature and linking it to the Islamic State by saying: “all of France is being menaced by Islamic fundamentalist terrorism.” Further evidence for this explicit link being made by the President was that he also used his televised speech to announce “the strengthening of [...] the actions in Syria and Iraq” (Mestre, Revault d’Allonnes and Bissuel, 2016). On July 16, ISIS, via its news agency, the Amaq News Agency, claimed responsibility for the attack by stating that a “soldier of the Islamic State” executed “a new, special operation using a truck”, and warning that “no
matter how much they enforce their security measures and procedures, it will not stop the mujahideen from striking.” (France24, 2016; Williams, 2016)

To evaluate the impact of this attack on the number of followers, we subsetted our dataset to only those observations that are observed more than once and fall within the window of 15 days before and after the attack, as we did for the Brussels attack. In this case, however, our period of analysis stops a few days after the attack, on 20th July 2016, and, consequently, we have a 15-day pre-treatment period and a 7-day post-treatment period. Given that the bandwidths that we will use are narrower than this time window, this will not affect our estimates.

Table 4, in Panel B, presents the main estimates of the impact of the terrorist attack in Nice on the number of followers in the accounts linked to ISIS.\textsuperscript{18} As we did for the Brussels bombing, our favorite specification is reported in column (3) because it uses the bandwidth suggested by Imbens and Kalyanaraman (2012). Under this specification, the Nice attack decreases the number of followers by 40 followers, which is equivalent to a 15% decrease from the mean number of followers throughout the eight-day period – four days to the left and four days to the right of the attack’s timeline. Additionally, the effect is consistently negative and statistically significant across bandwidths – the IK’s bandwidth, its half, or its double –, although the effect decreases as we incorporate data points that are further apart from the attack date. Interestingly, if we focus on the difference between the date before the attack and the attack date (bandwidth equals 1), the effect is also significantly negative. Figure 5b illustrates the discontinuous jump downwards that occurs between the date before and after the attack. This is consistent with the general pattern of a causal and negative effect of attacks on the number of followers in accounts associated with ISIS.

\textsuperscript{18}To avoid repetition, this estimates follow the same logic as Panel A for the case of Brussels.
5.1.2 All data: A panel study with individual fixed effects

We have just shown that the two major ISIS-related terrorist attack led to a substantive decrease in the online support of ISIS. In this section, the aim is to evaluate the same hypothesis as in the previous section; this time regarding the whole dataset which includes all major ISIS-related terrorist attacks during the period from March to July 2016.

Table 3 reports the impact of the number of victims in ISIS attacks on the number of followers of ISIS on Twitter on the day of the attack and in its immediate aftermath. Panels in the Table show the results across different discount rates. Models reported in odd columns, with no user-account fixed effects, are OLS regressions with clustered standard errors at the account level. Models in even columns are varying-intercept multilevel models where day-intercepts are modeled as a function of attacks perpetrated by ISIS. The set of models in the first column shows that the effect of the number of victims in the US and Europe is negatively related to the log of the number of followers of Twitter-related accounts.

However, these models may be biased due to distinct sample composition in the pre-attack and the post-attack periods. Thus, the second set of models in columns 3 and 6 focus exclusively on the within-account variation. This specification is closest to identifying a causal effect because it incorporates an intercept for each account and, consequently, controls for all between-account heterogeneity that could account for the relationship. In a consistent manner, we see that the effect remains negative and statistically significant.

If these models are capturing a true causal effect of bombing, we should expect that the effect of an extra victim in the United States and Europe should be stronger than the effect of a victim elsewhere. Consistent with our theoretical expectations, models in columns 3 through 6 also show a negative effect on the number of followers, although less consistently so and with smaller coefficients. In particular, the effect of an individual death in Europe and the US leads to a decrease in the log of followers in that same day of -0.0005 or -0.004, yet the impact of a victim outside the US and Europe is -0.0001 or -0.00008, that is, 5
times smaller. In other words, a terrorist attack outside Europe and the US requires about 5 deaths to achieve the equivalent effect to a single death within the US or Europe on the number of followers.

Altogether, together with the results from the formal section ((1) and (2) on page 18), these results empirically support the thesis that the external discouraging effect from the attacks dominates its external mobilization effect since we are able to detect a significant negative effect in the number of followers in the aftermath of terrorist attacks. While the results are statistically significant, we now turn to discuss whether the magnitude of the effect is substantive. Certainly, the large size of our sample allows us to detect effects that are very small in magnitude. Therefore, it is pertinent to discuss whether our parameter of interest captures a meaningful causal effect.

Before moving to the discussion of the magnitude of the effect, we should bear in mind that the estimated parameter does not capture the full impact of the discouraging strategy, but only the net effect of the discouraging effect after subtracting it from a potential mobilization effect. If we assume that both effects may be strictly positive in absolute values, then the estimated discouraging effect should be interpreted as a lower bound of its true effect.

Figure 4 reports the predicted value of the number of followers in ISIS-related Twitter accounts as a function of the number of deaths associated with a particular day in the sample. The number of followers in Twitter accounts related to ISIS is predicted to be close to 137 if there has not been an attack on that day nor on the dates immediately following. However, an attack in the US or on European soil would imply a decrease in the number of followers in that day and in its aftermath, and this decrease depends on the intensity of and the proximity to the attack. Approximately, each 25 victims leads to an average decrease of 2 followers in their accounts. By contrast, the magnitude of the decrease in the number of followers is much weaker because it requires an attack with 100 deaths to generate a change in the predicted value of 1 follower less in ISIS’s Twitter accounts.
In the interpretation of these results, we should take into account that the variable “number of deaths” does not truly account for the number of victims in a given day because the variable is measured in a way that allows for lingering effects of a given terrorist attack over time. Thus, to explore the actual impact of a discrete attack in the number of followers, we should assess not only the effect it has on a particular day, but the cumulative effect it has throughout the period in the aftermath of the attack. To illustrate the shifts in the predicted values as a function of the number of deaths in an attack, Table 5 reports the five most bloody attacks within our period of study, the date they occurred, their location, their death toll, and the predicted change in the number of followers caused by an attack alone.

The values in the last column are the predicted cumulative percentage change in the number of followers as a result of actual death toll related to attacks observed within our period of study. As a reminder, this means that the effect of the Brussels bombing had a value of 36 on 22 March 2016, 24 on March 23, 11 on March 24, 3 on March 25, 1 on March 26, and values indistinguishable from 0 thereafter. Hence, the simulation exercise involves summing these actual values in the variables produced by single attacks to obtain the change in the predicted value expected from each discrete attack. The results show that the largest impact was generated by the case of the attack in Nice with an average accumulative change of a 7.5% decrease in the number of followers. Another important impact is caused by an expected 6.6% decrease in the number of followers that would be expected from the Karrada bombings, a coordinated attack that killed 341 people on 3 July 2016 in Baghdad, Iraq. Even though not so important in magnitude, the attack in Istanbul and Brussels killing 44 and 36 people would be expected to cause a 3.8% and 3.2% decrease.

This simulation exercise allow us to illustrate that the magnitude of the effects observed in our models are significant statistically, but also quantitatively. To put these percentages in perspective, a 7% decrease from a starting point of 137 means an average of 9 fewer followers.
Taking into account that there are about 10,554 accounts related to ISIS throughout our period of study, this quantitatively means an expected drop of 94,986 in the absolute number of followings to ISIS. This is a substantial reduction in the size of ISIS’s Twitter audience.

### 5.2 Exploring the Mechanism: Demotivation or Deterrence?

In this section, we try to tease out the logic behind the negative effect of terrorist attacks on the number of followers on Twitter between the two potential explanations: the demotivation and deterrence effect. In particular, our evidence suggests that the demotivation effect constitutes a great proportion of the total negative effect found in our main results. We utilize the known geographic location of some accounts to test the deterrence effect. Hence, if the deterrence effect is the only driver of our main findings, then we should see that the negative effect is significantly larger among those accounts located in Europe and North-America, so closer to countries with stronger counter-terrorist agencies. Testing this means testing our third hypothesis. To approximate the geographic location of accounts, we use the accounts’ language to separate those accounts whose language is spoken only within Europe and North America from all other accounts.

To test the empirical prediction of the deterrence mechanism, Table 6 reports three model specifications with varying degrees of pooling across accounts, although results are not sensitive to the specification. Overall, the effect of terrorist attacks on the number of followers for all those who do not use an European-only language accounts ranges from $-0.00050$ (p-value < 0.001) (column 3: within-account variation) to $-0.0042$ (p-value < 0.001) (column 1: between-account variation). Yet, there is an important differential effect of attacks among those users that do use an European-only language. In particular, the effect of terrorist attacks among accounts that use an European-only language ranges from $-0.00011$ (p-value = 0.74) (column 3: within-account variation) to $-0.0017$ (p-value < 0.05).

---

19 This does not take into consideration that one user can be counted more than once because it may withdraw from following several posting accounts at the same time. Therefore, this should be interpreted as an upper bound estimate of the average decrease in the number of users.

20 see more details in the "Data" section above
(column 1: between-account variation). In other words, the effect of attacks to accounts with European-only language is about one fifth (22%), one third (35%), and two fifths (40%) of the effect on all other accounts. Though the difference in slopes is also significant for attacks outside Europe and the US, the magnitude of the overall effect of these makes problematic any substantive inference the first place (see above, Figure 4).

[Table 6 about here]

Overall, we see that the key prediction from the deterrence mechanism for that accounts closer to European countries should be more affected by attacks in European soil is not supported by the results. Thus, the finding that the negative effect of terrorist attacks among accounts with an European-only language is weaker than the effect on all other observations, which is contrary to the expectation based on the deterrence mechanism. This suggests that the discouraging rather than the deterrence mechanism is likely to be at work in our main analysis.

Conclusion

Twitter has become a crucial communication and recruitment tool for the Islamic State. This paper offers an innovative approach to learning about the characteristics of ISIS by tracking the reports made by Anonymous on their Twitter accounts. The new data allows our study to explore on an important question in political science in general, and the terrorism literature in particular: what are the consequences of terrorist attacks? While providing a full answer to this issue is beyond the scope of this research, we have provided new empirical evidence on the impact of terrorist attacks in the relationship between the organization and individuals.

Our results shed light on the dynamics between those on border between mild sympathizers - passive supporters - and those outside the organization. In particular, a strong empirical regularity has emerged from our analysis: terrorist attacks lead to a decrease in
the size of ISIS Twitter audience in the aftermath of an attack that is both statistically and substantively significant. These results speak directly to the literature of the effectiveness of violence for terrorist organizations, by showing that the discouraging effect among passive supporters dominates both the mobilization and the attentional effects that terrorist attacks have on them. Hence, if a terrorist organization seeks to broaden its base of explicit supporters through violent activities, we show that this strategy is counter-productive because not only does the Twitter audience not increase, but it significantly decreases.

Assuming rationality among terrorists, then the unaddressed puzzle is: why would terrorist organizations engage in violent activities if this causes a decrease in the number of people that will be reached or potentially recruited? We can speculate that the reduction in their Twitter audience is the price terrorists have to pay to accrue two benefits from attacks. First, attacks may not aim at broadening the base of supporters (external mobilization), but to increase the loyalty of those who were already members (internal mobilization); in other words, they might trade quantity for quality by radicalizing already-supporters at the expense of losing moderate supporters. Unfortunately, the internal mobilization dynamics is hidden given just the Twitter data. We do not see whether supporters move from passive to active, or from active to leader after terrorist attacks, because they are followers of the observed accounts both before and after an attack. However, potentially, this puzzle is definitely solvable. For example, one of the promising directions for the further research is to do a survey study by contacting some of the Twitter users who stopped following the ISIS-related accounts after the terrorist attacks.

Finally, while attacks backfire by reducing the size of the Twitter audience, they may aim at infusing terror in the civil population. This may benefit the organization in terms of exerting influence over foreign governments, as suggested by a body of scholarship (e.g., Bueno de Mesquita, 2005; Bueno de Mesquita and Dickson, 2007). Possibly, infusing terror among the population may increase the likelihood of governments to either provide policy concessions (Thomas, 2014), or provoke an excessive response of government for the advan-
tage of the organization (Bueno de Mesquita, 2005; Bueno de Mesquita and Dickson, 2007). Thus, the cost paid in a reduction in the number of people that can be reached through Twitter may be worth taking.

While this research focuses on a particular aspect of the relationship between the organization and society, we acknowledge that further research is required to fill the theoretical and empirical gaps that remain in this literature to fully understand the incentives of pursuing violent activities by terrorist organizations.

References


URL: goo.gl/nDOFuY


URL: goo.gl/QwAhHT


Ellyat, Holly. 2015. “‘Anonymous’ Hackers Declare War on ISIS in Video Message.”.


URL: goo.gl/6qKPzm


Holmes, Jack. 2015. “Anonymous Just Declared War on ISIS.”.


Mosbergen, Dominique. 2015. “Anonymous Declares War On ISIS After Paris Attacks.”.


Table 1: Dynamics across categories (individual-level dynamics)

<table>
<thead>
<tr>
<th></th>
<th>Mobilization</th>
<th>Discouragement</th>
</tr>
</thead>
<tbody>
<tr>
<td>External</td>
<td>Non-follower $\implies$ Passive supporter</td>
<td>Passive supporter $\implies$ Non-follower</td>
</tr>
<tr>
<td>Internal</td>
<td>Passive supported $\implies$ Active supporter;</td>
<td>Active supporter $\implies$ Passive supporter</td>
</tr>
<tr>
<td></td>
<td>Active supporter $\implies$ Leader</td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Individual category transition stochastic model

<table>
<thead>
<tr>
<th></th>
<th>leader group</th>
<th>active</th>
<th>passive</th>
<th>non-followers</th>
<th>observers</th>
</tr>
</thead>
<tbody>
<tr>
<td>leader group</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>active</td>
<td>(\alpha_0)</td>
<td>(\alpha_1)</td>
<td>(\alpha_2)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>passive</td>
<td>0</td>
<td>(\beta_0)</td>
<td>(\beta_1)</td>
<td>(\beta_2)</td>
<td>0</td>
</tr>
<tr>
<td>non-followers</td>
<td>0</td>
<td>0</td>
<td>(\gamma_0)</td>
<td>(\gamma_1)</td>
<td>(\gamma_2)</td>
</tr>
<tr>
<td>observers</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 3: The Impact of the Number of Victims in Attacks Linked to ISIS on their Number of Followers in Twitter

<table>
<thead>
<tr>
<th>PANEL A:</th>
<th>Dependent variable: Number of Followers (log scale)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Discount rate: 50%</td>
</tr>
<tr>
<td></td>
<td>(1.1) (1.2) (1.3) (1.4) (1.5) (1.6)</td>
</tr>
<tr>
<td>Number of victims (US &amp; Europe)</td>
<td>−0.002*** (0.0002)</td>
</tr>
<tr>
<td>Number of victims (outside US &amp; Europe)</td>
<td>−0.008*** (0.00006)</td>
</tr>
<tr>
<td>RE user-account</td>
<td>×</td>
</tr>
<tr>
<td>FE user-account</td>
<td>×</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PANEL B:</th>
<th>Discount rate: 75%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(2.1) (2.2) (2.3) (2.4) (2.5) (2.6)</td>
</tr>
<tr>
<td>Number of victims (US &amp; Europe)</td>
<td>−0.002*** (0.0002)</td>
</tr>
<tr>
<td>Number of victims (outside US &amp; Europe)</td>
<td>−0.0007*** (0.00007)</td>
</tr>
<tr>
<td>RE user-account</td>
<td>×</td>
</tr>
<tr>
<td>FE user-account</td>
<td>×</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PANEL C:</th>
<th>Discount rate: 100%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(3.1) (3.2) (3.3) (3.4) (3.5) (3.6)</td>
</tr>
<tr>
<td>Number of victims (US &amp; Europe)</td>
<td>−0.002*** (0.0002)</td>
</tr>
<tr>
<td>Number of victims (outside US &amp; Europe)</td>
<td>−0.0007*** (0.00006)</td>
</tr>
<tr>
<td>RE user-account</td>
<td>×</td>
</tr>
<tr>
<td>FE user-account</td>
<td>×</td>
</tr>
</tbody>
</table>

N total 239,434 239,434 218,727 239,434 239,434 218,727
N accounts 10,554 10,554 8,320 10,554 10,554 8,320
N days 127 127 127 127 127 127

Note: *p<0.1; **p<0.05; ***p<0.01. Constants omitted from the output. Columns 1 and 4: OLS regressions. Columns 2 and 5: Random effects model with varying-intercept at the level of user-account and day, with day-intercepts modeled as a function of attacks perpetrated by ISIS. Columns 3 and 6: Fixed-effects model at the account-level with panel corrected standard errors. The number of observations in the fixed effects models is lower because accounts with one observation are excluded from the dataset for identification purposes. In all models, we have excluded extreme observations in the dependent variable – the top 1% percentile, yet results are unaltered if we include them.
Table 4: Local Average Treatment Effects of Timing of the Attack on the Number of Twitter Followers

<table>
<thead>
<tr>
<th></th>
<th>PANEL A: Terrorist Attack in Brussels</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dependent Variable: Number of Followers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td></td>
</tr>
<tr>
<td>LATE: ( \bar{\tau} )</td>
<td>-25.0* (13.0)</td>
<td>-37.1* (16.3)</td>
<td>-19.8** (11.0)</td>
<td>-19.0** (9.5)</td>
<td></td>
</tr>
<tr>
<td>Bandwidth</td>
<td>1 day (half-IK)</td>
<td>3 days (IK)</td>
<td>6 days (IK)</td>
<td>14 days (double-IK)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>3,744</td>
<td>16,967</td>
<td>34,183</td>
<td>53,069</td>
<td></td>
</tr>
<tr>
<td>Pre-treatment N</td>
<td>2,858</td>
<td>8,542</td>
<td>17,077</td>
<td>17,077</td>
<td></td>
</tr>
<tr>
<td>Post-treatment N</td>
<td>2,825</td>
<td>8,425</td>
<td>17,106</td>
<td>35,992</td>
<td></td>
</tr>
</tbody>
</table>

|                  | PANEL B: Terrorist Attack in Nice |                  |                  |                  |                  |
|                  | Dependent Variable: Number of Followers |                  |                  |                  |                  |
|                  | (1)                              | (2)              | (3)              | (4)              |                  |
| LATE: \( \bar{\tau} \) | -58.6*** (14.1) | -127.3*** (22.4) | -40.1** (16.1) | -23.6** (11.4)    |                  |
| Bandwidth        | 1 day (half-IK)                   | 2 days (IK)      | 4 days (IK)      | 8 days (double-IK) |                  |
| Observations     | 3,744                             | 6,988            | 14,269           | 26,997           |                  |
| Pre-treatment N  | 2,003                             | 3,714            | 7,711            | 15,554           |                  |
| Post-treatment N | 1,741                             | 3,274            | 6,558            | 11,443           |                  |

Note: *\( p<0.1 \); **\( p<0.05 \); ***\( p<0.01 \). IK refers to Imbens and Kalyanaraman’s (2012) bandwidth.
Table 5: Predicted Percentage Change in the Number of Followers After Some Important Terrorist Events within our Sample Period

<table>
<thead>
<tr>
<th>Terrorist Attack</th>
<th>Date of the Attack</th>
<th>Location</th>
<th>Number of deaths</th>
<th>Predicted Cumulative % Change in the Number of Followers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brussels bombings</td>
<td>22 March 2016</td>
<td>Brussels, Belgium</td>
<td>36</td>
<td>-3.00%</td>
</tr>
<tr>
<td>Yemen bombings</td>
<td>23 May 2016</td>
<td>Aden, Yemen</td>
<td>45</td>
<td>-0.8%</td>
</tr>
<tr>
<td>Ataturk airport attack</td>
<td>28 June 2016</td>
<td>Istanbul, Turkey</td>
<td>44</td>
<td>-3.6%</td>
</tr>
<tr>
<td>Karrada bombings</td>
<td>3 July 2016</td>
<td>Baghdad, Iraq</td>
<td>341</td>
<td>-5.6%</td>
</tr>
<tr>
<td>Nice attack</td>
<td>14 July 2016</td>
<td>Nice, France</td>
<td>86</td>
<td>-7.1%</td>
</tr>
</tbody>
</table>

Note: Simulated values are based on the fixed-effects models from column 3 and 6 in Table 3.
Table 6: The Impact of the Number of Victims in Attacks Linked to ISIS on their Number of Followers in Twitter

<table>
<thead>
<tr>
<th></th>
<th>Number of Followers (log scale)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td></td>
</tr>
<tr>
<td>EUR-only language account</td>
<td>0.387***</td>
<td>0.069***</td>
<td>0.032**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.014)</td>
<td>(0.015)</td>
<td></td>
</tr>
<tr>
<td>Number of victims (US &amp; Europe)</td>
<td>-0.004***</td>
<td>-0.001***</td>
<td>-0.001***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.0002)</td>
<td>(0.00003)</td>
<td></td>
</tr>
<tr>
<td>Number of victims (outside US &amp; Europe)</td>
<td>-0.001***</td>
<td>-0.0001**</td>
<td>-0.0001***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.00005)</td>
<td>(0.00001)</td>
<td></td>
</tr>
<tr>
<td>EUR-only language account X Number of victims (US &amp; Europe)</td>
<td>0.002***</td>
<td>0.0004***</td>
<td>0.0004***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td></td>
</tr>
<tr>
<td>EUR-only language account X Number of victims (outside US &amp; Europe)</td>
<td>0.0002</td>
<td>0.0001***</td>
<td>0.0001***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.00002)</td>
<td>(0.00002)</td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RE user-account</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FE user-account</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N total</td>
<td>239,434</td>
<td>239,434</td>
<td>218,727</td>
</tr>
<tr>
<td>N accounts</td>
<td>10,554</td>
<td>10,554</td>
<td>8,320</td>
</tr>
<tr>
<td>N days</td>
<td>127</td>
<td>127</td>
<td>127</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01. Constants omitted from the output. Column 1: OLS regressions. Column 2: Random effects model with varying-intercept at the level of user-account and day, with day-intercepts modeled as a function of attacks perpetrated by ISIS. Columns 3: Fixed-effects model at the account-level with panel corrected standard errors with the number of victims interacted with a time-invariant characteristic of the accounts, language of the account. Model is column 3 is also referred as hybrid model because it provides variation in a time-invariant variable (language of the account) because of its interaction with a time-variant variable (number of victims), while maintaining any other characteristic of the account constant. Just as for the main analysis, the number of observations in the fixed effects models is lower because accounts with one observation are excluded from the dataset for identification purposes. In all models, we have excluded extreme observations in the dependent variable – the top 1% percentile, yet results are unaltered if we include them.
Figure 1: A Comparison of Google Trends to the Cumulative Lagged Measure of Deaths

Brussels Bombings (keywords “Brussels Bombing”)

![Graph showing the comparison between Google Trends and the cumulative lagged measure of deaths for Brussels Bombings.](image)
Figure 2: External processes (aggregate-level dynamics)

External Processes

\[
\begin{align*}
\text{Mobilization (+: Mobilization effects} \\
\quad \text{Attentional effects} \\
\text{Discouraging (-: Deterrence effects} \\
\quad \text{Demotivating effects}
\end{align*}
\]
Figure 3: Cumulative Lagged Measure of Deaths from Terrorist Attacks by ISIS (03/16/2016 – 07/20/2016)
Figure 4: Predicted Values of the Number of Followers Depending on the Number of Deaths and the Geographic Location of Terrorist Attacks
Figure 5: Interrupted Time Series of Followers by Two Major Attacks

(a) ISIS Attack in Brussels, Belgium (March 22, 2016)  (b) ISIS Attack in Nice, France (July 14, 2016)