Violence against civilians in the Syrian civil war

Andrew Halterman

4 July 2018

Abstract

Why have armed groups in the Syrian civil war deliberately killed so many civilians? Existing theories of civilian targeting in war offer indeterminate predictions about violence against civilians in civil war: targeting civilians can "drain the sea", but lose "hearts and minds", be rational or driven by emotions, carefully targeted or indiscriminately applied, or could be the inadvertent byproduct of conventional fighting. I compile the largest available micro-dataset on civilian death in civil war, comprising data on the dates, locations, and causes of over 100,000 civilian deaths in the Syrian war, along with fine-grained data on armed groups' territorial control and locations of arrests during the protests in 2011. Using this data, I systematically evaluate existing theories' abilities to explain violence in Syria. I find little support for prominent theories of violence against civilians that emphasize the importance of "hearts and minds," intelligence, and territorial control, principal-agent problems, or "desperation". Instead, strategic logics of deliberate civilian violence, especially "political" repression in areas of anti-regime mobilization and "war winning" mass violence explain the majority of casualties in Syria. The new micro-level dataset will contribute to other studies of violence in the context of civil war.

Word count: 9,286 (JCR: “Total word count, including everything, should be at a minimum about 8,000 and not exceed 11,000 words.”)
Introduction

Despite extensive research on violence against civilians in civil war, the literature presents mixed predictions about when, why, and to what extent armed groups in civil war will target civilians. Most practitioners of counterinsurgency see mass violence as encouraging defection to the other side. Violence creates cycles of revenge and hatred-driven opposition to the perpetrating side. Attacks against innocent people are morally repugnant.

Why, then, are civilians consistently victimized in civil wars? The ongoing Syrian civil war has seen widespread targeting of civilians during the conflict with half a million killed and the majority of civilians displaced.¹ What explains where and when violence against civilians occurs and who perpetrates it? Several competing theories have emerged to explain civilian killing, but little systematic work has directly compared their ability to explain violence across a single conflict. I systematically test three families of theories of violence against civilians in the Syrian civil war using a new micro-level dataset that offers the most detailed quantitative data on civilian casualties of any civil war, including the dates, geographic coordinates, and methods of civilian deaths, daily measures of degree of territorial control, and the number and locations of arrests in the early phases of the uprising in Syria.

Specifically, I test a family of theories concerned with identity and ideas, including principal-agent problems and emotions, a set of strategic theories about violence in the context of rational strategies for war-winning and political victory, and a “null” theory of collateral damage in frontline areas. Applying a set of existing theories to a single conflict, and specifically the civil war in Syria, offers several advantages. First, it is intrinsically important to provide an answer to the question of why armed groups in Syria are killing civilians. Second, the availability of detailed micro-level data on the conflict makes a systematic comparison of theories possible. Finally, making good-faith efforts to apply theo-

ries beyond their scope conditions gives us important information about when theories do not apply, but more importantly, the mechanisms or missing assumptions underlying their inability to generalize to new contexts.

I find little evidence for existing theories of territorial control and intelligence or principal-agent problems in explaining most casualties in Syria. Instead, most civilian casualties in Syria should be understood as an outgrowth of rational strategies of mass killing, specifically in the context of state repression and violence against areas of high anti-regime mobilization. The overwhelming number of civilian deaths in Syria are caused by the Syrian government, primarily as the result of indirect weapons employed against populated areas.

I argue that the theories that do not explain the conflict well do so primarily because their conception of the incentives of and choices available to actors to not match the conflict in Syria. Government forces have a wider range of heavy weapons than some theories predict, and strategies of collective targeting that some theories do not address. Finally, actors have a wider range of incentives than some theories account for, specifically concern with pre- and post-war political opposition to the regime.

Theories of violence against civilians in civil war

Remarkably, given the degree of attention that violence against civilians in civil war has received, no consensus exists in the literature on whether widespread violence against civilians should be expected. Many scholars and practitioners of civil conflict, especially those studying insurgency and counterinsurgency, see violence against civilians as irrational and counterproductive against civilians as harmful to the goal of winning the population's support (Nagl 2002; Nagl et al. 2008; Kocher, Pepinsky, and Kalyvas 2008; Lyall, Blair, and Imai 2013; Dell and Querubin 2016, 2017), leading to predictions that it would be uncommon. Other stands of research on both conventional wars and civil conflict recognize that violence against civilians can be motivated by emotions and identity, or that it may be a rational strategy in some cases.

I create a typology of the theories of armed group violence against civilians, categorizing them broadly
into a non-political “null” theory of neglect and collateral damage, theories that emphasize “ideas” and emotion, and a set of “strategic” theories that propose different reasons why targeting civilians is rational or irrational. Several theories of violence seem to be in operation at once in Syria. What distinguishes the accurate theories from inaccurate are the extent to which they account for the different capabilities and incentives of actors. Government and rebel forces have different wartime incentives, different access to weapons, and differing incentives for how the postwar order is to be structured.

**Null theory: collateral damage**

A simple “null” theory of violence against civilians would posit that violence is not deliberately targeted at civilians, but rather civilians are killed by accident or neglect in front line areas or areas of active fighting. This scenario includes both civilian casualties that result from “battlefield necessity” and those produced by carelessness, neglect, or malice in the context of ongoing military operations. This distinction is highly consequential morally (e.g. Walzer 1977) but less so politically. In either case, civilian are killed because they happen to be in an area of active fighting between armed belligerents. The observable implications of this theory are clear: civilians should not be killed away from frontline areas and rarely targeted directly.

**Ideas and emotions**

One set of theories of violence in civil war center on the role of “ideas” in conflict: ethnicity, individual desires for victimization, and the role of emotions such as revenge in spurring violence.

**Principal-agent problems**

The principal-agent literature on violence against civilians (e.g. Weinstein 2006; Shapiro 2013; Worsnop 2014) begins with an assumption about divergent preferences: low-level fighters have an appetite or preference for civilian victimization, and commanders prefer that their subordinates do not engage in violence against civilians. Under this account, violence is produced when principals are
unable to monitor and sanction fighters who engage in victimization. Principal-agent problems can arise from the adverse selection of recruits driven by a group's available resources, with recruits who are eager for immediate material payouts will engage in more civilian victimization (Weinstein 2006). Covert groups’ security-enhancing measures may prevent leaders from monitoring their subordinates (Shapiro 2013). Groups that lack centralized resources will not be able to punish unwanted behavior on the part of subordinates, and groups without intensive training will not be able to align preferences between commanders and soldiers (Worsnop 2014). Taken together, this suggests that civilian victimization will be limited in well-indoctrinated, overt, and consolidated armed groups.

If the principal-agent theory were in operation in Syria and explaining most civilian casualties, we would expect that governments forces should generally produce less victimization than non-state groups. We would expect that civilian victimization would be the result of low-level agents’ actions, not the actions of higher-up principals. We would therefore expect that most indiscriminate violence would therefore be committed by low-level units using small arms and light weapons that do not require higher authorization, and therefore little civilian victimization to result from heavy weapons, such as artillery or aircraft, that require higher authorization from the chain of command.

**Emotions**

Scholars of conflict have also pointed to emotions as an important cause of violence in civil war. Petersen (2002) offers clear predictions for when structural changes will create fear, hatred, resentment, and rage in the context of ethnic conflict, and which behaviors are likely to result from each. Balcells (2017) argues that the desire for revenge plays an important role in wartime violence, creating an “endogenous process” of violence (Kalyvas 2006).

**Ethnicity**

Ethnicity clearly plays a role in many civil conflicts, though the mechanism by which it results in violence against civilians is contested and clearly varies across cases. Some of this violence is “strategic”:
Posen (1993) suggests that civilians in an ethnic conflict increasingly identify as members of their identity and take steps to defend themselves. This produces a security dilemma, creating pressures for other ethnic groups to do the same and to create defensible territorial units. Other violence is motivated by emotions in the context of ethnicity (see above). Weidmann (2011) distinguishes between “top down” ethnic violence based on territorial concerns such as Posen (1993) or Toft (2005), and “bottom up” violence committed by individually motivated perpetrators. If ethnic violence were driving causalities in Syria, we would observe efforts to create ethnically homogeneous areas in a “top down” approach, and individual violence motivated by ethnicity in “bottom up” scenarios. The latter form of violence will always be more difficult to measure, given its dependency on the perpetrator’s state of mind.

**Strategic theories**

A large body of research focuses on the rational reasons why targeting civilians would be beneficial or harmful to armed groups.

**Hearts and minds/intelligence and territorial control**

Perhaps the most prevalent “strategic” theories of violence against civilians in civil war see it as counterproductive and therefore irrational. Both scholars (e.g. Kalyvas 2006; Berman and Matanock 2015) and practitioners (Nagl 2002; FM 3-24, Kilcullen 2010) have proposed a theory of armed group success in civil war that relies on the cooperation of the local population. If the “fundamental problem” (Kalyvas 2006) of irregular conflict is identifying enemies and enemy collaborators, civilian cooperation is crucial for gaining the intelligence needed to identify enemies. Cooperation is proved to armed groups that successfully control territory (Kalyvas 2006) or provide governance and services [Sunderland (1964); FM 3-24]. If winning the support of civilians becomes the *sine qua non* of counterinsurgency, violence against civilians, and especially indiscriminate violence, becomes counterproductive. If rational civilians are targeted indiscriminately, at best they will abstain from cooperation from the side committing violence, and at worst will support the other side for protection. (Civilians motivated by emotions will have their own reasons for opposing a side that has employed violence against them.)
Kalyvas (2006) makes more specific predictions about selective violence: it should peak in areas under partial control, where armed groups have enough presence to generate tips from civilians, but still the desire to consolidate control.

This set of theories was developed to explain violence in guerrilla insurgencies, not conventional civil wars with defined front lines. In conventional civil wars, the identification problem is reduced and armed groups have consolidated rears, meaning that civilian intelligence and logistical support is less important. The role of civilian support in conventional civil wars should not be discounted completely, however. As in conventional interstate wars, civilians produce the arms, food, and supplies needed by conventional intrastate combatants to sustain their fighting. This reliance on civilians implies that the same dynamics as guerrilla war could be portable to conventional civil war.

If the “hearts and minds”/intelligence/territorial control theory does apply to Syria, violence against civilians should be fairly uncommon and when present, specifically directed at individual civilians. Indiscriminate violence against civilians will push them toward the opposing side, making it irrational and therefore rare. Where violence does occur, it should be in areas of partial control, rather than fully contested or fully controlled areas, as armed groups attempt to consolidate their control.

**Strategic violence in wartime**

A literature on conventional interstate wars suggests that combatants may have reasons to rationally target civilians in wartime (e.g. Valentino, Huth, and Croco 2006) and militaries have adopted strategies of bombing populations, most notably during the strategic bombing campaigns of the Second World War. Pape (1996) provides the canonical distinction between “denial” strategies that target civilians to reduce their ability to produce arms, recruits, food, and supplies used by the combatants, and “punishment” strategies to impose costs on civilian populations and thereby pressure combatants to give up. Later work argues that a denial strategy could be at play in guerrilla wars to deny access to resources (Eck and Hultman 2007; Downes 2007; Fjelde and Hultman 2014; Koren and Bagozzi 2017) and also in conventional civil wars (Krcmaric 2018). A “denial and punishment” theory would predict greatest levels of violence in areas of Syria completely held by one side, especially areas with high perceived
support for opponents or areas where opponents generate a large portion of their matériel.

One of the important insights of the strategic violence literature is that civilians can be targeted “collectively”, not just indiscriminately or selectively (Sullivan 2012). Collective violence is directed at areas or groups where there’s some reason to believe that civilians support the opposing side, but not information on specific individuals to target. Adding collective targeting as a strategy available to combatants means that they can engage in violence with less need for intelligence.

A closely related “desperation” theory (Hultman 2007; Downes 2008) suggests that governments or rebels target civilians during protracted conflict or when they are performing poorly on the battlefield, in order to pressure the opposing side into conceding. As the war drags on, armed groups become more willing to try strategies they would not have considered at the beginning of the war and may be increasingly willing to trade long term costs (in civilian opposition) for short term benefits (in reduced will or ability). In Syria, this theory would predict that as the threat to the regime increases or the duration of the war increases, the number of civilians killed by the armed groups (especially the government) would increase.

**Strategic political violence**

Balcells (2017) suggests that governments experience not only wartime incentives to target civilians, but also broader political calculations that take into account the postwar order. “Indirect” violence against civilians, such as airstrikes on cities, are rational in areas with high pre-war anti-regime political mobilization. Civilians in these areas are seen as threats to the regime, both during the war, when they provide support for the rebels, and after the war, when they will pose political problems for the postwar order. Balcells, in studying the Spanish Civil War, used pre-war opposition vote share, finding greatest violence near a 50% tipping point. In traveling to Syria, the model would need to rely on alternative measures of political opposition, and the authoritarian logic of repression might result in a different tipping point than the 50% produced by electoral systems. Thus, the theory would predict greater violence, specifically indirect violence, in areas of high anti-regime opposition before the war.
The civil war in Syria

The civil war in Syria began as a series of protests in mid-March 2011, as part of a broader wave of anti-regime protests in the Middle East. The protests spread through April and the government response quickly became violent with security forces firing on protesters. By July 2011, a nascent Free Syrian Army began armed opposition to the state. Fighting intensified through mid-2012 and 2013, with rebel forces, including Jabhat al-Nusra, making important territorial gains. On August 21, 2013, a chemical attack in Ghouta led to several hundred deaths and drew greater international attention to the conflict. In 2014, fighting began between ISIS and other rebel groups, and US airstrikes began in September 2014.

The years 2015 and 2016 are especially relevant for this study, given the data availability I describe in the next section. In the first half of 2015, Nusrah and other mostly Islamic forces were advancing in Idlib and northwest Syria. Meanwhile, ISIS was advancing in Palmyra, Kobane, and al-Haskakeh. Russian airstrikes began in October 2015. Syrian government forces took Latakia in January 2016 and parts of Daraa Governorate. A ceasefire was in effect from February to July 2016, but territorial changes continued. Turkish military forces and Kurdish PYD and YPG forces each took territory in northern and northeastern Syria in the fall, and in December 2016, government forces took the last areas of rebel-controlled Aleppo.

The civil war in Syria shares important similarities with previous civil wars. First, as a “conventional” civil war characterized by high rebel military capacity and an absence of guerrilla insurgent tactics, the war is different from many of the wars studied in the civil war literature (Iraq, Afghanistan, Colombia, El Salvador, Chechnya, etc.) but is part of a growing population of conventional civil wars (Kalyvas and Balcells 2010). It has involved a great deal of foreign involvement, a growing feature of civil wars (Anderson 2016).

The war in Syria is also made different from others studied previously by its high prevalence of conventional, as opposed to guerrilla fighting. This prevents some theories from being applied without modification (see previous section). The war is also different in its availability of micro-level data.
Syria was a middle income country with high literacy rates and connection to the rest of the Middle East. Unlike civil wars fought in poorer countries, the war in Syria takes place in a context of ubiquitous phones and internet, making detailed local information on violence and armed actors much more available than in other civil wars. The presence of the Islamic State in Syria has also mobilized research and interest well beyond the attention paid to most civil wars, leading to more research and data on the war. These factors mean that researchers can begin to study the war even before it is over, hopefully providing better information to policymakers than they had in previous wars.

**Data and Methods**

To understand which of these theories, if any, explain the broad patterns of violence in Syria, I collect a new set of large-scale quantitative data on violence against civilians, along with information needed to test each theory. The use of large quantitative data sacrifices the detailed understanding of specific acts of violence that ethnographic or qualitative research can provide, but gains generality in return. The sum of civilian deaths in Syria results from a mixture of processes: some civilians will be killed in crossfire from a battle, others targeted after a denouncement, some killed by low-level units disobeying orders, and some by actors’ deliberate strategic violence. A useful theory will be able to account for as many of these deaths as possible.

I collect several types of data: the locations and dates of civilian deaths, the cause of death, groups’ control of territory, the degree of territorial control, distance to the front line, a measure of pre-war mobilization, and a measure of threat to the regime’s survival. I collect daily data on each of these and geocode them, when applicable, to the neighborhood, village, or city. I compile this information from a number of sources, enhance it using methods described below, and combine it into a single consolidated panel dataset, covering around 6,000 unique locations observed daily in Syria between 2011 to 2016 with the number and cause of civilian deaths at each point, and with territorial control available after 2015. The dataset thus consists of day-resolution counts of deaths in Syria, augmented with the geographic coordinates of the death at the neighborhood or village level, the group occupying
that location on that day, and its occupier and degree of control.

**Civilian deaths**

To collect information on civilian deaths in the Syrian civil war I use data provided by the Syrian Shuhada site. The Shuhada (“Martyrs”) dataset records the name, date, combatant status, cause of death, location of birth, and location of death for people killed in the Syrian civil war. The data covers from the beginning of the conflict in 2011 through the middle of 2016. The majority of the data, 65.7%, is drawn from the Center for Documentation of Violations in Syria (VDC), which compiles information on the death of civilians, rebels, and regime forces in Syria. Price, Gohdes, and Ball (2014) find the VDC dataset to be an accurate source with as good of coverage as any other single dataset. Another 23.8% comes from the Syrian Center for Human Rights, and the remaining 10.5% are compiled from news sources, YouTube, opposition groups, Facebook posts, and Local Coordinating Councils. The Shuhada dataset consolidates data from across these sources, standardizes their formats, and augments them with greater detail on the locations of Syrians’ deaths. The requirement that casualties be documented with a name, date, and location of birth and death suggests that the deaths included in the Shuhada and VDC datasets are likely to be underreports of the total. Their cooperation with opposition groups could also induce an underreporting of civilians killed in government areas or by rebel combatants. I scrape and format this dataset, merging information from separate English, Arabic, and combatant status pages into a consolidated dataset.

**Geocoding civilian death data**

The Shuhada dataset includes fine-grained geographic information on the location of civilian causalities, reporting the governorate and city of death for 96% of casualties, and reports an additional neighborhood of death for 26.0% of the casualties. Although this information is present, it has not been used by researchers, to my knowledge, because the locations are provided as free text, not as geographic coordinates. Any attempt to merge this dataset with other geographic datasets or to account

---

2Available at http://syrianshuhada.com/. A complete copy of the original dataset is available in the replication materials.
for distance in modeling requires that the text fields be converted to geographic coordinates. There are 4,543 unique Arabic-language place names reported by the Shuhada dataset, and 4,507 unique English transliterated place names. Looking up geographic coordinates for each of these locations by hand would be a difficult undertaking. Instead, I create an algorithm that can take in place names in either Arabic or English and return their geographic coordinates, using information about the governorate and type of place (city vs. neighborhood) to return a precise match. This language-agnostic algorithm can be used anywhere that researchers have structured geographic place name data and would like to produce geographic coordinates. Automating the task makes the process much faster, reproducible, easily extensible to other datasets, and allows the dataset to be rapidly expanded as more data is made available.

In brief, I look up each place name from the Shuhada dataset in a search database populated with the Geonames gazetteer (Wick and Boutreux 2011) of place names, limiting the search to the given governorate in Syria and using a series of rules to determine the best match from the potentially multiple results. I attempt the search first using the Arabic place names; if no results are found, I attempt again with the transliterated English place names. After completing this process, I was able to obtain geographic coordinates for 90.0% of the civilian casualties: 104,134 out of a total of 116,026 civilian deaths recorded in the dataset.

**Direct and Indirect Causes of Death**

Some theories make specific predictions about the types of lethal violence directed against civilians, namely whether the violence is selective/direct or indiscriminate/indirect, with direct violence being violence committed with small arms, and indirect violence being the result of air strikes, artillery fire, and in the case of Syria, chemical weapons. Using Balcells’ definitions of direct and indirect violence, I map each of the reported causes of death in the Shuhada dataset to those two categories. I also code regime airstrikes separately, given their importance in Balcells’ theory.

---

3 Assuming a rapid 60 seconds per lookup, this would take approximately 150 hours of work.

4 Further details are provided in the supporting information.
Territorial Control

Understanding the role of territorial control in explaining civilian casualties requires information on which groups controlled which locations on each day. Territorial control maps in Syria are very popular and several organizations, including the Institute for the Study of War, Caerus Associates, and the New York Times have produced control maps since the beginning of the conflict. While very useful for getting a sense of the conflict from a macro level, these maps do not have the right spatial and temporal resolution to be useful.

I use a dataset compiled by the Carter Center, which tracks changes in territorial control at the village or neighborhood level from January 1, 2015 to present.⁵ I scraped and formatted the data into a location–day panel dataset of territorial control. Each coding of a territorial control is accompanied by a citation to a news report or social media report describing the capture, making the data much more transparent and trustworthy than it would be without this information. The major drawback of the Carter Center data, in contrast to the more traditional maps, is that their coverage only extends back to January 1, 2015, while the civilian casualty death data from Shuhada extends back to the beginning of the conflict in 2011. After cleaning, the territorial control dataset contains 5,676 towns/neighborhoods, each of which has data on which group controls it, coded on a daily basis.

The Carter Center’s data treats control as a hard indicator: every location is controlled by exactly one group in equal degree. To adapt this data to theories concerned with degree of control, I develop and compute two new continuous measures of territorial control. The first measure is the distance to the nearest locale controlled by a different group, to reflect how close an area is to the front line or nearest enemy. A second measure is the proportion of locales in a place’s immediate neighborhood of 15 that are controlled by other groups, to better capture the “precariousness” of control.⁶ Being surrounded by “enemy” areas is a conventional analogue to the degree of control measure that Kalyvas uses. Being surrounded makes loss of control more likely and the presence of opposing forces more

⁵The project is described at https://www.cartercenter.org/peace/conflict_resolution/syria-conflict-resolution.html, and the map is available at https://d3svb6mundity5.cloudfront.net/dashboard/index.html
⁶A neighborhood size of 15 was chosen qualitatively by examining average distances to neighbors and inspection of maps across urban and rural areas.
salient. If civilian cooperation is necessary to produce direct violence and civilians fear retaliation if control should flip, we should expect less direct violence in highly precarious areas.

I also compute and use as a control a measure of how urban or rural an area is, given the known issues with ignoring different urban and rural dynamics in conflict (Kalyvas 2004; Douglass and Harkness 2018). Settlements included in the Carter Center dataset are much denser in urban areas than rural. I calculate the median distance to these nearest 15 locales as a measure of settlement density.

In an irregular war of the type Kalyvas mostly considers, front lines are permeable and civilians need to fear that groups have read into their opponents areas where they could punish civilian collaborators. The mechanism is different here. Front lines are relatively impermeable, but the lines can shift over time. These areas of partial control are insecure in that that are most at risk of flipping control, triggering the retaliatory killings that are a day-to-day fear in irregular civil wars.

I join the civilian casualty and territorial control datasets by making a panel of all Carter Center locations and assigning civilian casualties to the nearest location. After matching locations in the causality dataset and the territorial control dataset, most locations are within 1 kilometer of their match (see the supporting information). Only 232 deaths are more than 5 kilometers from a locale in the territorial control dataset.

Arrests

A key contribution in Balcells (2017) is to consider the role of political opposition to the government before the onset of war and the implications this opposition has for violence during the war and the post-war political order. In her main case of the Spanish civil war, she uses pre-war vote share as a measure of opposition to the regime. Syria’s rigged elections make Syrian electoral data useless for this purpose. As an alternative measure of the government’s perception of anti-regime mobilization, I compile geolocated data on arrests of activists and protesters in 2011. During the initial protest phase

---

7 See the supporting information for details on the distance algorithm used.
8 In the 2007 presidential election, the last before the onset of war, only 0.18% of votes were against Bashar al-Assad, on 96% turnout, according to the International Foundation for Electoral Systems. http://www.electionguide.org/elections/id/110/
of the conflict, arrests by government forces are a good proxy of government perception of opposition to the government in an area. Through July of 2011, when the Free Syrian Army began operating, the Syrian regime had free access to the entire country. The Syrian Center for Statistics and Research, an NGO, collects data on civilian causalities in the war, along with arrests of civilians by the government.⁹ This data includes the name of the arrestee and the date and location of the arrest. Following a similar procedure to the Shuhada dataset, I scrape and geocode the arrests in the dataset to produce geographic coordinates for each arrest. After linking each location name to its geographic coordinates, I then merge it with the data on locations and civilian causalities during the rest of the conflict. Rather than assigning the arrests only the closest location in the panel data, I use a distance-decay function to assign partial credit to other nearby locales to account for spillover across locations.¹⁰

Regime threat

As the situation for one side becomes more “desperate”, it may be more willing to turn to indiscriminate violence against civilians to stave off defeat. To measure threat to the regime, I employ contemporaneous forecasts about the probability of Assad leaving power made as part of the US intelligence community’s Good Judgement Project. The Good Judgement Project (GJP) was a project sponsored by the Intelligence Advanced Research Projects Activity to study whether accurate geopolitical forecasts could be produced from crowdsourced decisions. Since the conclusion of the project, all forecasts have been publicly released to researchers. While the precise performance of the project compared to classified analyses are not publicly available, the popular reporting on it suggests that its forecasts matched or exceeded the intelligence community’s existing approaches.¹¹ The overall accuracy of the GJP forecasts suggest that their specific forecasts about the prospects of the Assad regime may be the best available assessment available to actors at the time of the probability of Assad remaining in power.

---

⁹More information and the original dataset are available at https://csr-sy.org/. A copy of the scraped dataset is available in the replication materials.

¹⁰I use the decay function \( f(n, d) = n \cdot \frac{1}{1 + \left( \frac{d}{2000} \right)^2} \), where \( n \) is the number of arrests and \( d \) is the distance in meters. An area within 500 meters gets full credit for each arrest, an area 2 kilometers away receives 0.5, and an area 10 km away receives about 1% of each arrest. More details on the function and parameters are provided in the supporting information.

As his probability of retaining power increases, a “desperation” theory would predict increasing violence against civilians. I take the questions asking about whether Assad will remain in power on a specific date in the future, adjust these forecasts to account for the time left in the forecast period, and add this information to the larger dataset.12

Summary Table

The data sources I compiled are presented in the table below, along with their dates of coverage. I combine all of these datasets into a location-week panel dataset. For each of the 5,676 unique locations, I construct a weekly panel with each variable. I also create an indicator variable for the occurrence of any direct or indirect deaths in a locale-week: \( I \{ \text{deaths}_{it}^{\text{direct}} > 0 \} \), \( I \{ \text{deaths}_{it}^{\text{indirect}} > 0 \} \). Binarizing the data is justified both theoretically and empirically. The differing casualties from single direct and indirect attacks are difficult to compare, as indirect attacks will generally produce more. Moreover, most theories of violence against civilians are concerned with whether, where, and how the attacks are occurring, not with the deadliness of any particular attack. Finally, the modal weekly casualty count, given that a casualty occurs, is 1.

Table 1: Summary table of datasets and sources

<table>
<thead>
<tr>
<th></th>
<th>Data Source</th>
<th>Modifications</th>
<th>Dates Covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cause of Death</td>
<td>Shuhada</td>
<td>Balcells’ direct/indirect coding</td>
<td>2011–2016</td>
</tr>
<tr>
<td>Cause of Death (alternative)</td>
<td>VDC</td>
<td>Balcells’ direct/indirect coding</td>
<td>2011–2018</td>
</tr>
<tr>
<td>Location of Death</td>
<td>Shuhada</td>
<td>Geocoding algorithm</td>
<td>2011–2016</td>
</tr>
<tr>
<td>Distance to Frontline and Enemy Areas</td>
<td>Carter Center</td>
<td>Geocoding algorithm, nearest neighbors, distance calculation</td>
<td>2015–2016</td>
</tr>
</tbody>
</table>

12The precise questions and aggregation strategy are available in the supplemental materials.
<table>
<thead>
<tr>
<th>Data Source Modifications</th>
<th>Dates Covered</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Early Phase Arrests</strong></td>
<td>Geocoding algorithm, nearest neighbors, distance calculation</td>
</tr>
<tr>
<td>Syrian Center for Statistics and Research</td>
<td></td>
</tr>
<tr>
<td><strong>Threat to Regime</strong></td>
<td>Forecast duration adjustment</td>
</tr>
<tr>
<td>Good Judgement Project forecasts</td>
<td></td>
</tr>
</tbody>
</table>

**Results and Discussion**

Equipped with a cleaned and geocoded panel dataset on the locations and causes of civilian deaths and our measure of which villages are controlled to what extent by whom, we can begin to adjudicate between competing explanations for civilian death in Syria. The first major finding is that two of the greatest three causes of death, air strikes and artillery, are employed almost entirely by the government (Figure 1). The Syrian Observatory for Human Rights, a Syrian opposition NGOs, claim that approximately 75% of the casualties in the war have been caused by government forces.\(^{13}\) Another NGO, the Syrian Network for Human Rights, claims approximately 90% of the civilian causalities of the war were killed by government forces.\(^{14}\) Even if these numbers are high, it seems clear that the majority of deaths are caused by government forces, making their behavior of highest importance to explain.

**Principal-agent theories**

Figure 1 reports the breakdown of causes of death in the Syrian civil war by cause over time from the beginning in 2011 to mid-2016, using both the Shuhada dataset and as a second source the VDC data. The first obvious finding of this table is that a large percentage of civilian deaths recorded in this dataset, 41%, are caused by artillery and warplanes that only the Syrian government or outside states employ and that are not selective by any measure.\(^{15}\)

\(^{13}\)http://www.syria.hr/en/?p=108723
\(^{14}\)http://sn4hr.org/blog/2018/09/24/civilian-death-toll/
\(^{15}\)The proportion is higher when using the Violations Documentation Center data, which extends longer into the conflict.
Looking at how the cause of civilian death changes over time, an interesting pattern immediately emerges: two of the major causes of death, shooting and shelling, have tapered off over time. The cause of death that has seen the greatest increase since 2014 the conflict is aerial bombardment, which is perhaps the least selective or “direct” cause of death. The greatest threats to civilians in Syria, according to this data, is the instruments of conventional armies, directed at civilians.

The extensive use of air strikes and the tremendous loss of civilian life resulting from air strikes (around 250–500 deaths per month; the counts are higher in the alternative VDC dataset), indicate that the Syrian government has decided that causing civilian deaths is either not important or is actively good for their cause.

Using the data I have, I can only test the coarsest predictions of the principal-agent theories, namely that groups engage in different rates of violence against civilians. I cannot directly test the mechanism of violence as the product of a breakdown in hierarchy without detailed qualitative research. That said, the types of deaths produced in Syria can shed some light on the theory’s predictions.

First, for the government at least, a large amount of violence is directed from the top, in contradiction
to the predictions of the theory. Multi-year shelling and airstrikes against civilians are not the result of poorly-indoctrinated rebels engaging in violence their superiors would not want them to. The violence, in the case of the government with its attributable and indiscriminate weapons, is clearly sanctioned from the top.\textsuperscript{16}

\textbf{Intelligence and Territorial Control}

Territorial control–intelligence or “hearts and minds” theories make specific predictions about selection and indiscriminate violence and territorial control that are not borne out in the data on Syria. The first evidence against the theory is the prevalence of indirect violence in the data. The theory sees indiscriminate violence as irrational and counterproductive, but we see from Figure 1 that it is very prevalent. Figure 2 shows the predicted probabilities of a locale-week experiencing at least one death of either direct or indirect cause, as a function of changing distance to the nearest “enemy” locale. The left panel initially indicates a Kalyvasian relationship between civilian deaths and a measure of territorial control. Violence in highly “precarious” areas, those surrounded by opposing groups, is very rare. Likewise, violence is rare in fully controlled areas with no unfriendly neighbors. Several complications present themselves, however. First, the relationships are not statistically significant.\textsuperscript{17} Moreover, the patterns in the left panel do not match the theoretical expectation. Indirect violence should be high in contested areas with (hypothesized) low information. In fact, direct violence is much more likely in nearly-surrounded areas. Indirect violence peaks in areas with one-third to one-half of their neighbors being controlled by other groups. Second, once controlling for distance to the front line and other control variables, the relationship disappears (right panel). Direct violence is higher with more “enemy” neighbors and tapers off as areas become more homogeneously controlled.

The relationship between both forms of violence and distance to the “enemy” is remarkably strong (Figure 3). The closer a locale is to the nearest place controlled by another group, the higher both se-

\textsuperscript{16}A comparison across rebel groups suggests that the role of ideology in violence against civilians has not been fully explored. Kurdish forces may use their discipline to refrain from killing civilians, while ISIS, because of its beliefs about what is appropriate and effective, may use its discipline to kill civilians, but in ways and for reasons that advance their ideological objectives.

\textsuperscript{17}See supporting information.
Figure 2: Relationship between civilian killing and proportion of “friendly” locales in the immediate neighborhood of 15. The model with controls includes distance to nearest “enemy” locale, the distance squared, the distance adjusted by the average distance to nearby locales, and a measure of urbanness. The relationship is not significant after accounting for clustered errors at the locale level. Computed with n = 44,521 civilian casualties from 2015 and 2016. See the supporting information for more details.
Figure 3: Relationship between civilian killing and distance to the nearest enemy-held area. Predicted probabilities of a locale-week experiencing a civilian casualty by varying distance to the nearest enemy area from four logit models. The models are for direct/indirect deaths and models with and without control variables. The no controls model includes only distance to the nearest locale controlled by a different group, and its squared term. The model with controls includes the distance adjusted by the average distance to nearby locales to account for urban/rural differences, and the fraction of nearby areas controlled by the same group, plus squared terms of each. Computed with n = 44,521 civilian casualties from 2015 and 2016. 95% confidence intervals. See the supporting information for more details.

Selective and indiscriminate violence is. In 2015–2016 period, 69.5% of deaths occur within 5 kilometers of an area held by a different group. The intelligence–territorial control theory sees selective and indiscriminate violence as imperfect substitutes. Groups prefer selective violence, but sometimes resort to indiscriminate violence when attacking enemy areas. In Syria, the two forms of violence go alongside one another, potentially indicating greater complementarity than the theory assumes.

Selective violence occur more in contested zones on the front line in Syria, where the theory predicts “no” [sic] selective violence (Kalyvas 2006, 204), than it does in areas of partial control where the theory predicts it will be greatest. In absolute numbers, recognizably selective violence like field executions and kidnappings are a small portion of civilian deaths in Syria. Even if the theory were correct for the subset of deaths that are selective, the data makes it clear that this form of deliberate, intimate killing is a small fraction of the puzzle of civilian deaths in Syria. The theory does correctly predict that most
deaths will occur in the border areas between armed groups, but this prediction is not unique to the theory and therefore does not greatly bolster its credibility.

There are several possible reasons for why the theory does not match the data from Syria. First, the mechanisms generalized from the theory may not travel from guerrilla wars to conventional civil wars. Without the ability to access contested areas for punishment, information about collaborators is useless to combatants and selective violence is not possible or helpful. Second, it could be that the theory operates, but only for a small set of deaths, and that civilians are simply likely to die where fighting is occurring or because they are in an area that has been deliberately targeted by armed actors for violence against presumed enemy civilians. Finally, the evidence could indicate problems with two key assumptions of the theory. Selective or direct violence may not be the joint product of civilian cooperation and armed groups, and instead the sole product of armed groups operating without intelligence or collecting intelligence without relying on local collaborators. Second, the evidence could indicate that armed groups’ preferences for mass violence against civilians is much higher than the theory sees as rational.

Strategic violence

The set of theories that fare the best are the ones that see deliberate killing of civilians as potentially effective, especially the “political” theory of Balcells (2017). Strategic mass violence theories expect much more indirect, deliberate, and indiscriminate violence than other theories, and this matches the evidence from Syria well. Figure 1 shows a large amount of indirect violence. It also shows how early in the war government forces resorted to indirect means of killing. This seems to indicate an early political strategy of violence, in contrast to the “war-winning” strategies of mass violence that predict its emergence after long, stalemated or threatening insurgencies (Valentino 2000; Valentino, Huth, and Balch-Lindsay 2004; Downes 2007).

Much of the violence against civilians in Syria is clearly organized from the highest levels of the military and increased between 2012 and 2015 as the rebels threatened the survival of the government. Aerial weapons, used almost exclusively by the government and outside parties, and artillery, which is
Figure 4: Predicted probabilities of a locale-week experiencing a civilian casualty with varying 2011 arrests. Four logit models, for direct/indirect deaths and models with and without control variables. The “arrests only” model includes only 2011 distance-decayed arrests and its squared term. The distance decay gives neighboring areas partial credit for nearby arrests depending on the distance from the arrest. Mean weighted arrests = 7.79; median = 0.036. Forty-four areas have more than 200 adjusted arrests and are not shown. The model with controls includes the distance to nearest enemy-controlled locale, the distance adjusted by the average distance to nearby locales to account for urban/rural differences, and the fraction of nearby areas controlled by the same group, plus squared terms. 95% confidence intervals accounting for clustering at the locale-level. See the supporting information for more details, including the corresponding regression table.
Figure 5: Side-by-side comparison of 2011 arrests locations and indirect violence during the war. The upper panel shows the locations of arrests in 2011. Grey dots show locations with no arrests. Blue dots have at least one arrest and are sized according to the number. The lower map shows the locations of indirect violence. Locations with indirect deaths are indicated in red, with the size of dot representing the logged number of deaths.
predominantly used by the government, are two of the greatest causes of deaths in this dataset and are not being used solely against enemy combatants on the battlefield.

If Balcells is right, we would expect greater indirect (indiscriminate) violence in areas of opposition to the regime in the beginning phases of the war. This data shows a remarkably strong relationship between arrests in 2011 and both forms of violence in subsequent years. Areas with no (adjusted) 2011 arrests had extremely low probabilities of experiencing at least one causality per week in subsequent years. This probability rises rapidly and exponentially as the number of arrests increases. The relationship between arrests and violence against civilians is almost identical for direct and indirect violence. Balcells’ theory would expect greater indiscriminate violence in areas with more pre-war anti-regime mobilization. This data indeed shows this relationship, but also finds that the probability of an area experiencing direct violence increases even more steeply with arrests. Second, this data shows a convex relationship: each additional arrest predicts more violence than each previous arrest. Balcells expects a diminishing relationship between pre-war opposition and violence, however. The different shapes of these curves may indicate the difference between democratic and authoritarian strategies of elimination. In Spain, only the median citizen needed to be a supporter. In Syria, the incentive is to eliminate all opposition.

Figure 5 shows the relationship between arrests and violence in a more qualitative way. Inspecting the results in this way helps reveal differences between the predicted and actual locations of violence. First, eastern Idlib Governorate and eastern Aleppo Governorate have both seen more indirect violence than 2011 arrests would predict. These could be the consequence of “war-winning” strategic violence, indirect violence directed against rebel-held areas as part of military operations in the area.

Latakia, on the northwestern coast of Syria, saw some early uprisings during the protest phase and a strong government crackdown, both in the city (marked on map) and in its southern countryside. During the war itself, however, the governorate has been relatively peaceful and supportive of the regime. The dot to the northwest of Damascus is Al-Zabadani, the first city taken by the Free Syrian Army in January 2012. Control was traded back and forth three more times between 2012 and 2017.

Theories of strategic mass killing, especially Balcells’ theory with its emphasis on pre-war mobiliza-
Figure 6: Duration-adjusted forecasts of Assad leaving power as a measure of regime threat, compared to the number of total and “indirect” (i.e. bombing and shelling) civilian casualties per day in Syria. The “desperation” or regime threat theory predicts a positive relationship between regime threat and violence, especially indiscriminate, against civilians. A loess line shows little relationship, however.

tion, very accurately describe the patterns of violence in Syria. Violence is concentrated in front-line areas and indiscriminate violence, including bombing, is concentrated in areas that saw high pre-war mobilization.

Regime threat

The regime threat hypothesis suggests that increasing threat to the regime should produce greater violence against civilians. The data does not find such a relationship, however. Figure 6 shows no relationship between threat to the regime, as measured by Good Judgement Project forecasts, and violence against civilians. The slope is indistinguishable from zero and its sign changes with minor adjustments, such as different daily lags. This figure provides evidence that threats to the regime do not seem to be the primary predictor of levels of violence.
Collateral Damage

With the quantitative data alone, absent a measure of where military forces are undertaking offensives, I cannot rule out a collateral damage explanation. A collateral damage theory would expect violence in front line areas, but concentrated especially in areas undergoing major military operations. It would not explain violence directed deliberately at civilians. A host of qualitative evidence from Syria indicates that violence is indeed directed deliberately at civilians and in areas that are not undergoing military combat at the time.

Conclusion

Ultimately, civilian deaths in Syria have been and will continue to be produced by a mixture of different processes. No single theory will be “correct” for all civilian causalities in Syria. The observed data from Syria, however, accords better with strategic theories: civilians are killed deliberately and indiscriminately, either as “enemy” civilians to be targeted directly, or as inconsequential causalities killed in the course of military operations. A small portion of deaths are selective, but their pattern does not match the predictions of the most prominent theory on selective violence and territorial control.

Applying theories of guerrilla war beyond their scope produces useful information on why they break down in conventional contexts. Combatants in conventional civil wars have a wider range of actions available to them than in irregular conflicts. Defined front lines make it possible to collectively target “enemy” civilians, and heavy weapons and air power make long range attacks against civilians possible. Authoritarian governments come to civil wars with a unique set of capabilities and incentives. As strong, security-oriented states, they often possess strong security forces and the intelligence and control to use them effectively. As incentives, they have a two-part logic of mass violence: an “eliminationist” strategy to kill or displace opponents of the regime, and a deterrence strategy to prevent future uprisings against the state through brutality. The large effect of pre-war arrests on subsequent targeting is evidence that the regime is using the war to eliminate areas of opposition to its rule, in the way Balcells describes. In contrast to Spain, where the concern was in shaping the median citizen after the
war, in Syria, there was no expectation of post-war democracy. Rather than ensuring that opponents
are in the minority, an authoritarian government may seek to eliminate as many as possible. Viewing
the violence as repression also accounts for the use of weapons: direct and indirect weapons are used
very similarly, and in a similarly indiscriminate fashion.

Brutal violence against civilians may also serve the function of deterring future mobilization against
the regime. The early violent repression of 2011 and the continuing regime violence against populated
areas can be seen in the context of a long history of violent repression, including most famously, the
massacre in Hama in 1982. Seeing chemical weapons in a logic of deterrence also helps explain their
use. Relatively ineffective for battlefield use and carrying high risks of international punishment, they
nevertheless create terror that could be used to deter future uprisings.

These conclusions have important implications for scholarship on killing in civil war. First, it high-
lights the important of meso theories of civil war killing that account for politics. Much of the existing
work on civil war and killing in civil war has operated either at a high-level structural level (e.g., the
large-n work of the early 2000s) or at the very micro-level, focusing on individual motivations and
emotions. Second, it points to the need for better theories of regime type in civil war killing. Authori-
tarian states face different incentives than democracies or third-party governs in how they wage civil
war. Regimes have incentives to use civil wars to reshape the post-war political order (Balcells 2017),
but authoritarian states may operate under stronger incentives than democracies to eliminate their
opponents.

Methodologically, the paper makes several contributions. The data provided here can inform future
studies of Syria: the automated techniques for geocoding text and measuring territorial control make
extensions of this work easy to implement. Neither method is specific to Syria or Arabic, meaning
researchers should be able to take them to other conflicts. Finally, the geocoding techniques are appli-
cable well beyond civil war. The same techniques can be used to create better data and understanding
on a host of behaviors, including protests, electoral violence, and government service provision.

The field on civilian victimization in civil war is expansive and growing. Testing existing theories on
new cases is a crucial component of advancing our understanding and reducing our overconfidence
about the causes of violence against civilians in civil war. This study takes a step toward doing so, and, through its methodological contributions, will make it easier for others to contribute in this literature and others.
References


