UNIVERSITY OF CALIFORNIA SAN DIEGO

ESSAYS ON EXTREMIST MEDIA STRATEGIES AND RADICALIZATION

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Political Science

by

Gregoire Loren Phillips

Committee in Charge:

Professor Barbara F Walter, Chair
Professor Claire Adida
Professor Eli Berman
Professor David Lake
Professor Margaret Roberts
Professor Branislav Slantchev

2022

The dissertation of Gregoire Phillips is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

University of California San Diego

2022

DEDICATION

This dissertation is dedicated to, and would not have been possible without, all of the people that held me up along my long, winding journey to this degree. First and foremost, I dedicate this dissertation to my partner Michael Kershow, without whom I could not have made it through the dark twists and turns of the last few years of plague and global turmoil. You are my light at the end of even the darkest tunnels. I also could not have gotten here without the continued support of my cohort-mates. Andres Gannon, Rachel Schoner, Luke Sanford, Mariana Carvolho, Kathryn Baragwanath, Lucas De Abreu Maia, Christina Cottiero, Brian Engelsma, Zoe Nemerever, Todd Levinson, and Mackenzie Rice: you each carried me through this journey – through our late nights and seemingly endless journey together through classes, defenses, conferences, revisions, and life changes. In addition to my peers, I also would be remiss if I didn't thank the faculty of the Department of Political Science at UC San Diego, the School of Global Policy and Strategy, and the Department of Economics for their support – most especially, my committee members Barbara F. Walter, David Lake, Claire Adida, Branislav Slantchev, Molly Roberts, and Eli Berman. Each of you taught me something that brought me closer to this milestone, and this dissertation exists in large part because of your efforts throughout this process. Last but certainly not least, I dedicate this dissertation to my family, both chosen and biological. Mohammad Salamah, Alex Velarde, Meghan Wynne, Morgan Funderburk, Corinne and William Wolfram: you all represent just a few of the people that lifted me up emotionally during the darkest moments of the last six years. I also would not be here without the life and wholeness given to me by Steven and Juliet Phillips, Nichole and William Martinez, Delinda Otto, Cathryn Fair, and the rest of my sprawling, loving family support system. This, too, is for you.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

IS	Islamic State
ISIS	Islamic State in Iraq and Syria (al-Sham)
AQ	Al Qaeda
AQAP	Al Qaeda in the Arabian Peninsula
PCA	Principal Component Analysis
OLS	Ordinary Least Squares
STM	Structural Topical Model
ML	Machine Learning

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

The following summarizes my contribution to each paper in this dissertation (introduction and conclusion are entirely my own).

Chapter 2, Measuring Message Strategies in Extremist Media, is a solo authored paper. This paper has its origins in my final paper assignment for Molly Roberts' Text As Data course, and I humbly acknowledge Prof. Barbara Walter's, Prof. Dotan Haim's, and Prof. Molly Robert's contributions to this work through their employment, collaboration in very early drafts, and feedback/teaching respectively. For that project, I worked with Dotan Haim to conduct simple unsupervised and unsupervised content classification tasks on a smaller subset (2 years) of extremist propaganda data. We conducted these analyses on data I had collected in 2015-2016 as a research assistant for Prof. Barbara Walter, along with supplemental data collected following the end of that RA work collected by me.

This analysis differed from the paper presented here in three critical ways. First, the scope of the data used in the analysis was far more limited; whereas we used only text data in the form of titles for the analysis in the course paper, I take advantage of both the transcribed words spoken in videos, audio releases and texts, and the identified objects identified in text, in this paper. Second, the sophistication of methods employed is greater: I approached this paper with a far more theoretically grounded set of categories corresponding to the theoretical heart of this dissertation (Indoctrination, Co-optation, and Intimidation) rather than the far earlier appeals brainstormed for the term paper, and used an expanded scope of machine learning methods for my supervised learning analysis. Third, this paper was written entirely on my own, without contributions by any of the above in the writing process.

Chapter 3, Mainstreaming Extremism, is also solo authored. This paper originated from a recurring question I had while working with Prof. Barbara Walter on our project investigating which extremist groups use online propaganda in Iraq and Syria: given that a group decides to adopt online propaganda as a tool, what strategic logic might extremists employ when determining what kind of content to disseminate to which groups of people? This paper represents my best answer to this question. I humbly acknowledge the employment and mentorship/collaborative support of Prof. Barbara Walter in collecting the data used in the empirical analysis of this paper, as well as the countless hours spent in her office digging deeper, and deeper, and yet deeper into how to properly articulate the research question and dig down into the theory of the paper. I also humbly acknowledge Prof. David Lake's contribution through a similarly long stretch of office hours probing the three core strategies that emerge at the heart of this paper. Finally, I acknowledge the intellectual contribution of Prof. Tamar Mitts, whose feedback throughout our time as collaborators was instrumental in shaping my thinking on this topic.

Chapter 4, Propaganda and Radicalization in an Internet Age, is co-authored. This paper is the product of a collaboration between Tamar Mitts, Barbara Walter, and me that began when Tamar approached us about the possibility of matching her data on social media radicalization and our data on online media produced and disseminated by the Islamic State (that I had begun collecting as a graduate student researcher with Prof. Barbara Walter). This paper then became the most extraordinary – and extraordinarily rewarding – collaborative venture of my time in grad school. To say working with Prof. Mitts and Prof. Walter is life-changing would be an egregious understatement. I will forever be grateful for their mentorship and guidance during our time working on this paper together.

Given that this piece is co-authored, it is perhaps most important for me to establish my contributions to the paper. I contributed to this paper in four ways:

1. I collected the online media data used to track disseminations of Islamic State propaganda on Twitter. This was primarily done as a GSR while working with Prof. Walter.
2. I contributed to our understanding of the theoretical drivers of extremist online engagement that serve as the exposition and background to the paper. Over the course of several drafts of the paper, I contributed significantly to our working understanding of the appeals used in propaganda from the perspective of the literature on radicalization and extremism in civil war.
3. I contributed significantly to the empirical design of this paper – most actively in the brainstorming and study design phase. In the early stages of this paper, Prof. Mitts, Prof. Walter and I were at a crossroads in determining how we both classify media releases into meaningful content categories and empirically show the effect of each content category on expressed preferences and behavior on social media. I brought the idea of reducing the words spoken in the video to transcriptions, while Tamar added that we could use object recognition to identify the objects that appear to enhance our detection of potential violence. This piece of the empirical design then served as the foundation for our strategy of disseminating tracing, which tracked disseminations of media scoring high on thematic content categories and traced the changes in online behavior across groups of individuals matched on pre-exposure expressions of support for the Islamic State who were exposed to media and those who were not.
4. I piloted early components of our empirical strategy, including verifying the usability and implementation of Amazon Rekognition in detecting objects that could indicate the

presence of violent content in media. While many of the downstream empirical tasks were carried out by Professor Mitts and her graduate student researchers, I played a key role in piloting these methods before we finalized and proceeded to executing our empirical strategy.

Outside of these contributions, I acknowledge the extraordinary efforts carried out by Profs. Walter and Mitts that made this paper possible.

Finally, I would also like to acknowledge the immense amount of time, care, and dedication invested in me by the faculty in my department – and, among those not named above, Profs. Claire Adida and David Lake. This dissertation would not have been possible without their support through these long years, and I am forever grateful for their care.

VITA

2015 Bachelor of Arts, University of California, Los Angeles

2017 Masters of Arts, University of California San Diego

2022 Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Mitts, Tamar; Gregoire Phillips, Barbara F. Walter. "Studying the Impact of ISIS Propaganda Campaigns." Journal of Politics, 2022 https://doi.org/10.1086/716281

FIELDS OF STUDY

Major Field: Political Science

Studies in International and Comparative Politics
Advisor: Professor Barbara F. Walter

ABSTRACT OF DISSERTATION

ESSAYS ON EXTREMIST MEDIA STRATEGIES AND RADICALIZATION

by

Gregoire Phillips

Doctor of Philosophy in Political Science

University of California San Diego 2022

Professor Barbara F. Walter, Chair

Why do extremists produce and disseminate such a wide, and often contradictory, variety of media? What is the relationship between extremist media and on-the-ground operations? This dissertation seeks to explain why extremists invest in diverse communications operations through the lens of strategic political communication. Building upon a foundation of discovery and taking the Islamic State's media operations as a case scope, I begin with an empirical dissection and

classification of Islamic State online propaganda disseminated from 2012 to 2017 – revealing a rich tapestry of variation in content and dissemination patterns across substantively different themes and media appeals. I then construct a theoretical framework for understanding the strategic logic of extremist media strategies that introduces three primary strategies of extremist media operations – indoctrination, co-optation, and intimidation – before exploiting changes in territorial control in the Islamic State's operational history to illustrate how changes in battlefield realities affect media strategy, finding a significant shift away from strategies of indoctrination and toward strategies of co-optation and (to a more limited extent) intimidation. Finally, I explore the consequences of extremist media on online follower behavior in coauthored work demonstrating a strong, significant, and in some cases counter-intuitive, effect of extremist media on online expressions of support for the Islamic State, a desire to travel to Islamic State battlegrounds, and Anti-western sentiment. Taken together, these contributions represent the most comprehensive empirical and theoretical treatments of extremist media in the literature to date, and serve as a foundation for future scholarship on the strategic use of media by non-state actors moving forward.

1. Introduction

> *"I hear everyone saying they're going to cut our heads off for not growing a beard. I'll tell you what… Saddam wanted to choke me with gas. Al-Maliki wouldn't care if my children starved to death. And America lumps me in with the rest of these people anyway. Say what you want about their violence, I don't think it's so good, but at least [the Islamic State] gets things done for my people. That's a hell of a lot more than the rest of them can say. I'll grow a beard and say my prayers if they help me feed my kids."* Conversation with an Iraqi ex-patriot cab driver in Jordan, July, 2014

1.1 The Puzzle

July 29, 2014 was a day like any other in Amman, Jordan. The sun rose high above a hazy sky alive with the sound of three million people rushing to start their workday. As an undergraduate research scholar way out of his depth doing field work for the first time in a country halfway across the globe, I was just starting to get a feel for the rhythm of the city. While Amman, as the capital and commercial center of Jordan, is quintessentially Jordanian, it only takes a few days in the city to notice just how Jordan's location at the nexus of a region wrought by three of the deadliest civil wars in modern history has influenced the daily ins and outs of life on its streets. The streets are alive with Syrian, Palestinian, and Iraqi Arabic conversation and music, and the cab drivers frequently dangle the flags of their foreign homelands from fastening of their rear-view mirrors. As I stepped into a cab that morning, I hardly noticed the miniature Iraqi flag adorning the dashboard in front of me, even as its owner, 'Ali, greeted me in Arabic tinged with a thick Iraqi accent. I wouldn't know it for several years, but my short time with 'Ali would shape my trajectory as a scholar and fundamentally change the way I thought – and still think – about extremist influence in an internet age.

After a few greetings and pleasantries, ‘Ali wasted no time asking me where I was from. As a pale, blond man with a backpack and clear expression of bewilderment at the seemingly lawless rules of the road (there were definitely rules – I just didn’t yet know them), I stuck out among the locals. After a brief conversation about California, university life in America, and my research plans in Jordan, I pivoted the conversation to him, asking how he came to work in Amman. He smiled wryly before switching into English. “You’re American – you don’t want to know about me… I’m the enemy.”

I was not entirely surprised. I had been given several versions of this answer before; why would a young American possibly care to get to know people that their country so frequently labeled as terrorists, militants, and extremists? Nevertheless, no one before had described themselves as “the enemy.” I was intrigued.

Cautious but captivated, I dug deeper, insisting that I did care and was interested in his story. After several refrains, he relented, telling me the story of his family’s struggle to get by in their homeland of Northern Iraq. Most of his family came from Mosul, the largest city in the northern province of Ninawa situated adjacent to the autonomous region of Kurdistan. They had fled to from Mosul to Baghdad during the American invasion in the early 2000s hoping to avoid much of the militant violence taking over the post-Saddam north but were forced back to Mosul by the increasingly bloody sectarian violence that gripped Baghdad under its first post-war Shi’a prime minister, Nouri al-Maliki. He described a life lived constantly between destitution and the barrel of a gun; no matter where his family moved, they were in danger. All they wanted was security and continuity, and they weren’t very much interested in who provided it to them. But all they got was violence and uncertainty. He described a rough life – one that culminated in his

family's move to Amman to escape escalating violence in 2009 during the Sunni Awakening movement.

"This all sounds very difficult," I said, moved by his story. "But I am still confused about one thing: what makes you the enemy?" At this, his expression changed from pleasant to cold and unreadable. He reached for the cell phone on the dashboard that he was using to navigate me deep into the heart city and swiped up to reveal a background photo of a series of black-masked militia men marching on a dirt road. They could have been anyone, but the flag blowing in the wind behind them, with its tight white calligraphy and a rough oval contrasting a stark black background, would have been recognizable anywhere: these were fighters for the Islamic State. "I want them to win," he said. "I see what they are doing out there on the internet. They're gonna free us. They're gonna make my country a better place."

It would be easy at this point in the story to label 'Ali an extremist. The problem isn't that he was, but rather that he *wasn't.* Throughout his story, and the explanation that followed his revelation as an Islamic State sympathizer, he stressed several times that he didn't agree with their interpretation of Islam or the harsh tactics that they used against journalists and foreign soldiers. He didn't much like my country or people like me, but he didn't want to harm me in any way. Rather, his assessment of and support for the Islamic State came from a place of harsh practicality: the previous status quo, as he perceived it, was steeped in a dictatorial past of oppression and a sectarian present of exclusion. From what he saw of the Islamic State following their recent takeover of several large, urban areas in Iraq and Syria, they were actively improving a system that he perceived to be working against him from the start. He liked what he saw.

But what exactly did ‘Ali see? How exactly did a cab driver in Jordan have enough information on how the Islamic State behaved on the ground in places like Mosul to embrace their cause despite disagreeing with them on virtually every ideological dimension? The answer is simple: he watched it all happen online. When I asked ‘Ali how he knew all of this was really happening, he said he had seen it on his friends’ social media. The picture that he had shown me? It was a still from a video he had picked up on Twitter shared by a family member. He may have joked about how much more competent his three-year-old was with his mobile phone, but he knew enough to access extremist content online and make ready use of that information when making decisions about which group to lend his support in his homeland.

This answer to this question could seem obvious to the nearly 70% of adults worldwide that own a cellphone with an internet connection, but the implications of the enormous expansion of telecommunications technology and its reach into even the most active conflict zones have eluded much of the contemporary research on the local dynamics civil conflict. Most of the existing literature focuses on the use of these technologies to radicalize and indoctrinate abroad – an important phenomenon, but one that ignores the effects of extremist efforts to broadcast and compete effectively at a local level. We’ve taken major steps as researchers to study the effect of ICT elsewhere but have largely ignored its effect on the local dynamics of influence in conflict – much to our own peril on the ground.

Why? We have struggled to understand how even the most radical extremist groups are able to compete so effectively beyond their ranks of ideological adherents because we have failed to recognize how extremist groups use precisely the same tools that they use to radicalize the sympathetic abroad to attract broader support at the local level. Extremists don’t win on the ground by just dismantling or defeating the state, but rather by convincing people that they are reliable,

just, and viable alternatives to the state – and while the production and dissemination of media doesn't directly lead to a larger following for such groups, it *does* enhance their ability to send messages to larger, more diverse audiences than they otherwise would be able to reach. Extremists might stay alive by hiding in the mountains or blending in among local populations, but winning the hearts and minds of people in an increasingly connected world where people rely on state-provided goods and services means that extremist groups stand to gain access to tremendous human and capital resources by coming out of hiding and competing with governments in the open.

Advances in ICT have, in effect, served to amplify the messaging strategies of groups traditionally limited to the fringe of influence. Extremist groups across the ideological spectrum have taken advantage of this – expanding the reach of their messaging through radio, television, and internet channels to a breadth of audiences previously inaccessible to all but those mustering the resources and support of the state. Understanding how extremists adapt these messaging strategies to shifting goals and priorities on the ground to tap into this support is paramount to redirecting support from these groups in the long run and crafting more sustainable peacebuilding policies that address the weaknesses in current counter-insurgency strategies.

1.2 Mainstreaming Extremist

At first glance, it's safe to say that most people expect extremist media to be dominated by indoctrinating ideological messages – pushing its consumers to accept extremist beliefs and radicalizing them for their cause. Particularly in the context of Islamic extremism, these beliefs are reinforced by scholarship which seeks to study the radicalizing effects of the most violent of media releases, focusing people's attention on the intimidation tactics seeking to draw people in with

success on the battlefield (Berger & Stern, 2015), religious claims to righteousness among followers (Maher, 2017), and the role of violence in the radicalization process (Mitts et al., 2020).

But extremist messaging isn't always as it seems. In fact, according to recent data (Walter & Phillips, 2020), if we were to put all the videos produced by the Islamic State into a bag and pick one out at random, we would only have a 1 in 3 chance of pulling out a video featuring extreme violence, and we'd have less than a 1 in 6 chance of pulling out a video on radical Islamist teachings. Extremist leaders certainly hold and espouse radical beliefs, but we wouldn't always know it from the messages they disseminate in their media.

This is puzzling because so much of what we know about violent extremism tells us that these groups succeed by doubling down on their radical beliefs and radicalizing vulnerable people to their cause (Berger & Stern, 2015; Walter, 2017). However, when we dig down into the communities within which some of these groups operate, we see a radically different picture of community bargaining, compromise, and negotiation that translates the diverse needs of society and the marketplace into functions that embed these groups within local settings (Ahmed, 2018). Likewise, when we dig into the messaging strategies of groups like the Islamic State, we are more likely to find a video of soldiers paving roads, directing traffic, or paying public employees than we are to find a formal sermon on the virtue of killing infidels. If we spend long enough looking, we are likely to find both among a veritable catalogue of highly varied messaging content. What explains the substantial variation in extremist messaging? What strategies do extremists actually pursue when crafting their messaging to the broader audiences with whom they seek to communicate when they seek to gain literal ground in the race for hearts and minds? And how do these different strategies resonate with their followers – and, perhaps most interestingly, their increasingly dominant followings online?

In this dissertation, I argue that extremists tailor their messaging strategies not simply to the extremist principles of their group identity, but rather to the shifting realities of their political and military organizations. While typically differentiated from traditional political actors and rebel groups by the influence of fringe ideological goals and the use of violence to pursue them, extremists remain fundamentally *political* actors subject to the same material, social, and infrastructural needs as groups that do not share their extreme doctrinal tendencies. The political nature of their enterprise manifests in the need to secure the support of enough individuals and material to accomplish their goals – whether those goals be toppling a foreign government, ushering in a global revolution, or as is more often the case, taking control over a certain territory and instituting their own set of preferred governance policies. Yet extremists are often put at an initial disadvantage in acquiring these things by the very thing that sets them apart from other political actors: their extremist identity. Extremists, by definition, assume an identity at the fringe of the ideological spectrum and develop reputations typically associated with violence toward those who do not do the same – acts that may very well screen potential recruits, donors, and supporters for favorable qualities (Walter, 2017) but that also limit the buy-in from an expanded audience from which they may want or need to draw support in the future.

When the time comes to attract support beyond the fringe, I argue that extremists turn to messaging to strategically manipulate their group narrative – their goals, methods, and constituency – in order to secure the support they need to succeed. Over the course of the three essays at the heart of this dissertation, I develop an empirical and theoretical framework of strategic messaging that starts from the needs and priorities of extremist organizations as they navigate the often-complex reality of their operations on the ground. The first essay constructs the empirical motivation and foundation enabling such a project – drawing from machine learning applications

to text and video data to create the rich empirical environment necessary to explore these questions with sufficient technical and generalizable rigor. The empirical frame of the first essay enables a theoretically grounded analysis of data typically relegated to low-frequency, predominantly qualitative analyses – expanding my horizons for exploration in essays 2 and 3. Drawing from existing historical and political science scholarship on propaganda, political messaging, and persuasion, I outline three broad messaging strategies in the second essay that extremists can pursue in order to secure support: indoctrination, co-optation, and intimidation. These three messaging strategies capture what I argue to be a theoretically rich representation of the strategic games that extremist groups may play with their followers – ranging from efforts to shift people's beliefs, to efforts to entice them with rewards, to threats to punish them with intimidating displays of brutality or violence – and I test how these messaging strategies vary across extremist territorial control to determine the extent to which extremist goals influence their media strategy. I then turn from the factors that influence extremist media strategy to its effect on online audiences in essay three in a collaborative essay alongside Tamar Mitts and Barbara Walter that explores how different appeals used in extremist media strategies influence the online behavior of their followers.

1.3 The Strategic Micro-foundations of Extremist Messaging Strategies

The primary theoretical contribution of this dissertation is a simple strategic framework of extremist messaging strategies. This framework first breaks down the micro-foundations of extremist messaging. I start with extremist goals and strategic needs: in order to understand the messages that extremists will seek to communicate to others, we need to understand what these groups want to accomplish, and what they will need – or think they will need – to do so. I then describe the prospective audiences from which extremists may draw from in acquiring the things

they need to accomplish their goals – sorting audiences along ideological and geographic dimensions to simplify our understanding of a complex process for the purpose of theory-building. From these three components – goals, needs, and potential audiences – I deduce three messaging strategies that extremists may pursue to convince their audience to provide them with the resources, support, and compliance necessary to accomplish their goals

Putting these components together, I identify three strategies that extremists seek to employ to attract support and compliance:

1. They can *indoctrinate* people – shift or exploit people's existing beliefs to make them subscribe to the same goals, thus motivating them to take up their cause;
2. They can *co-opt* people – entice people with promises of physical, political, or social rewards without shifting their ideological beliefs to conform to their goals;
3. They can *intimidate* people – force people into compliance with threats of physical or psychological punishment for interfering with or resisting their operations.

These strategies represent three simple, distinct games that extremists seek to play with potential supporters and detractors. On their own, these games give us a static snapshot of what extremists want. However, in context, these games can tell us a lot about the changing needs and strategies of extremist groups on the battlefield of hearts and minds. These games reflect the on-the-ground realities that extremists face: as extremist's goals, leadership, and fortunes on the battlefield change, so, too, do their needs from potential supporters.

While these games are distinct, we should not expect extremist groups to invest in any one strategy at a time. As extremist groups expand to communicate with different audiences, the messaging strategies they craft to convince those audiences to help them – advertently or not –

accomplish their goals may very well expand with them. We should therefore expect extremist messaging strategies to be diverse, at times, reflecting the multitude of interests of those with whom they seek to communicate to secure resources, recruits, acquiescence, or surrender. We should also expect extremists to exploit the geographic, technological, and linguistic differences among their audiences to more carefully target each group with its tailored messaging strategy – tailoring their local messaging strategy to the cultural, political, social, and linguistic characteristics of region to boost the persuasive appeal of their media. This dissertation most deeply engages with the geographic, technological, and ideological facets of these strategies. However, it is my hope that this work sparks a deeper exploration into how these appeals engage audiences across these social, cultural, technological, and ideological cleavages.

1.4 Empirical Application: Islamist Extremism and the Islamic State

In all three essays included in this dissertation, I apply this theoretical framework of extremist messaging to a mixed-methods empirical analysis of the local messaging production and dissemination of one organization, the Islamic State. I focus on the Islamic State for two reasons. First, groups like the Islamic State have generated a highly disproportionate amount of attention by Western governments and media in not only their media production, but also their territorial contestation and control, across the globe – providing an enormous amount of highly granular data on territorial extent and changing operational realities that are typically not available for extremist groups. This attention is no accident; the unprecedented fall, rise, and fall again of the Islamic State constitutes the culmination of a series of state and interventionist failures in MENA governments that laid the groundwork for extremist success (Feldman, 2014) while threatening many of the fragile political structures put into place by Western interveners during the first decade and a half of the new millennium. This heightened attention inspired and empowered scholars like

Aaron Zelin, Mara Revkin, Tamar Mitts, and so many others to collect the high frequency, highly granular data on the Islamic State that I used as the foundation and inspiration for the data that appears in this dissertation (Zelin 2014, 2016; Revkin, 2018, 2020; Mitts, 2020; Mitts, Phillips & Walter, 2022)

Second, the phenomenon of Islamist extremism itself contains a substantive, geographic, linguistic, and cross-national richness unrivaled in the modern world. Overshadowed in the past by the underground international communist movement, Islamist extremism emerged by the early 1990s as the primary militaristic manifestation of anti-statism and revolution across the Middle East, the African continent, and Central Asia and gained its most notorious foothold in the annals of history in the series of attacks on western targets best exemplified by those launched by al-Qaeda against the United States on September 11, 2001. Where many past studies of non-state messaging and propaganda have focused on the foremost extremist movements of their day, from the pressing threat of fascism and anarchism to the growing specter of communism, we turn today to Islamist and emerging right-wing extremism for modern parallels with rich enough data to support the analysis necessary to bring our theories to testing on the real world. It is my sincere hope that the substantive focus on Islamist extremist groups like the Islamic State combines the extraordinarily rich availability of data with the timely benefit of insight into what remains a pressing series of policy questions on how to understand these movements as they continue to exert influence worldwide.

1.5 Core Contributions

The core contributions of this dissertation are threefold. First, I introduce novel theoretical and empirical frameworks for conceptualizing messaging strategies as the culmination of strategic

games played by extremist groups with different audiences of potential supporters, detractors, and subjects of rule. This framework provides a lens through which to view and make sense of the often chaotic, multi-faceted thematic content produced and disseminated by groups hoping to blur the lines between their extreme ideological core and the material benefits of cooperation and acquiescence in their organization. As I break down in greater detail in chapters 2 and illustrate in chapters 3 and 4, this framework is flexible to several different breakdowns of thematic content and independent of the content categories selected by me in this research. Importantly, this framework's contribution lies in understanding the strategic logic of messaging strategies as tied not just to an ideological imperative, but to operational realities faced by groups competing with very real and present needs to accomplish their goals.

The framework, in short, is the following: extremist groups, like their state counterparts, are constantly engaging in several strategic actions to persuade, intimidate, or otherwise obtain the things they need to survive, thrive, and achieve their goals. These actions include material measures to obtain these things: the use of violence to push for their goals on the battlefield or intimidate others into giving them what they need, negotiations with competing and collaborating groups to combine and focus their efforts, the distribution of goods and services to their members and supporters to build a consistent base of support, and many, many other tools from the toolbox of success. However, these actions also include immaterial measures designed to shape individual's perceptions and expectations of the potential benefits of supporting an extremist group – and the potentially prohibitive costs of not doing so for themselves and others.

In motivating this framework, I argue that engaging only with extremists' tactics on the battlefield of guns, money, and negotiation misses a critical set of strategies that extremists employ to compete on the battlefield of hearts and minds. Namely, we lack a theoretically well-motivated

understanding of the types of strategies that extremists can employ to communicate the pros and cons – the carrots and sticks – of potential support of their cause to different audiences. Extremists don't just use media to draw attention to their material largesse, military victories, and thriving memberships (although they do attempt to do these things as well), but rather to shape the information environment and ethical, social, political, and economic calculus of supporting a group based on their expectations of what is right for them and their families. We understand how extremists use real-world violence and largesse to shift this calculus, but I argue that we lack an understanding of how rebels strategically employ messaging to leverage the information environments of their audiences in their favor.

My theoretical contribution seeks to fill this gap by investigating the role that three core messaging strategies – co-optation, indoctrination, and intimidation – play in the strategic relationship that extremists have with different audiences through their media outreach. Relying on decades of groundbreaking scholarship on the strategic motivations of non-state actors, I break down the motivation for rebels to engage with each of these strategies before further discussing the likely times, places, and audiences with which each strategy is likely to provide the greatest strategic benefit to extremist groups in Chapter 2.

Second, this dissertation introduces two new datasets that bring together several existing and new streams of data to provide scholars with an unprecedently clear picture of the operational scope and messaging diversity of the most militarily successful extremist group of our time: the Islamic State. Both datasets represent unique contributions to our existing knowledge. Together, each allows scholars to answer new and exciting questions about the structure, function, rise, and downfall of an organization that briefly controlled a territory the size of Greece with a population nearly three times that of Jordan. Beyond the Islamic State, these data provide a template for

understanding some extremist groups' relationship with territorial governance and administration, while the dataset on messaging strategies provides new data on the media releases and thematic content of two additional groups used as comparison and generalization examples in Chapter 6. These datasets form the empirical foundation of this dissertation, but beyond their usefulness here, it is my sincere hope that others use this data as a resource in future endeavors.

Finally, this dissertation presents the first comprehensive, high-volume mixed methods analysis of extremist propaganda to date. By bringing an unprecedented degree of granular, local messaging data to bear to address questions of messaging strategy, this dissertation departs from more traditional qualitative treatments of messaging strategy common among historians (Doob, 1938; Bob, 1984; Stanley, 2014, 2020) and expands upon the work of pioneering scholars like Molly Roberts (Roberts, 2019), Tamar Mitts (Mitts, 2019), and Mara Revkin (Revkin, 2019) seeking to bring the quantitative tools of the text-as-data revolution in political methodology to the forefront of the messaging research frontier alongside our qualitative toolset to understand how extremist groups seek to communicate strategically with a diverse host of audiences.

1.6 Roadmap of the Dissertation

This dissertation brings together three essays centered on understanding how extremist groups can use media strategically to communicate different messages to a diverse array of audiences at different times. These three essays largely draw from the same empirical source: propaganda and territorial control data on the Islamic State, collected between 2015 and 2019 and representing seven years of propaganda (2011-2018) produced and disseminated across at least thirteen countries in the Middle East, North Africa, Europe, and South Asia. While each essay approaches a different facet of the puzzle laid out at the beginning of this introduction, they each

represent a unique contribution that, when brought together, improve our collective understanding of extremist media strategy and its effect on online audiences.

The first essay, "Measuring Messaging Strategies in Extremist Media: A Content Analysis of Islamic State Propaganda," provides the empirical foundation for the quantitative study of extremist media strategies. At the heart of this essay, I introduce a methodological framework for analyzing online content produced by extremist groups online that employs both unsupervised and supervised machine learning methods to the analysis of large volumes text, audio, and video data. The goal of this framework is to allow researchers flexibility in using unsupervised methods to explore the potential underlying content dimensions their data while illustrating a key application of the more theoretically informed supervised methods to classify, label, or score their data in accordance with pre-defined characteristics. I then apply this framework to an analysis of Islamic State propaganda across eleven countries from 2011 to 2018 to illustrate how both unsupervised and supervised learning methods can help us better understand high-frequency text, audio, and video data in ways that better conform to theory-grounded quantitative research.

The second essay, "Mainstreaming Extremism: Indoctrination, Co-optation, and Intimidation on the Battlefield of Hearts and Minds," constructs a theoretical foundation for understanding how territorial ambitions affect the ways in which extremist groups use media to attract supporters and detract resistance in areas they seek to control. I propose three main hypotheses in the second essay of this work. First, I hypothesize that shifts in the an extremist group's operations on the ground from small-scale operations on local targets without territorial aspirations to large-scale operations seeking to contest and control territory lead to corresponding shifts in the group's messaging strategy – shifting extremists from a predominantly indoctrinating strategy of messaging seeking to draw in a small cadre of true believers dedicated to their

ideological cause to a wider net of co-optation strategies seeking to draw support or acquiescence from individuals now looking to the Islamic State as a governing power. Second, I hypothesize that shifts in operational strategy toward territorial control will fundamentally shift the way that extremist groups depict violence against non-combatants – shifting groups from depicting polarizing uses of less discriminant violence toward civilian groups broadly (*enthusiast* violence) that attract only the most devoted followers to uses of violence toward civilian groups that stress compliance (*deterrent* violence) with extremist authority. Last, I hypothesize that these shifts in extremist messaging strategies are "sticky": even when extremists shift away from strategies of territorial control, the effect of the shift on their messaging strategies endures well beyond the tenure of their actual administration. I test these hypotheses using media data coded into the content categories explored in the supervised learning section of essay one and find robust support for each hypothesis.

Last but certainly not least, I round out the three-essay trio with "Propaganda and Radicalization in an Internet Age", a collaboration with Tamar Mitts and Barbara Walter that examines the consequences of online media strategies on social media audiences through a study of content disseminations on Twitter. This essay brings together the empirical framework introduced in essay 1 and an extension of the theoretical logic introduced in essay 2 to explore the consequences of diverse media strategies employed by extremists on contemporary social media audiences. We bring these themes together to ask a simple question: what types of appeals increase support for extremists online? Using the Islamic State as our test case, we bring together thousands of online media releases with millions of tweets across five years of data to show that, contrary to popular expectations, violent content is far less effective at eliciting support from online followers – and is more likely to deter all but the most extreme supporters of the Islamic State's movement.

Instead, appeals corresponding to the broader umbrella strategies of indoctrination and co-optation are far more effective at increasing support across the board. We illustrate these empirical trends in great detail and point to the importance of future work in disentangling the mechanisms of media strategy from their effect on supporters.

Bringing this all together, I conclude the dissertation with an assessment of the contributions and policy implications of this work. If extremist messaging is, in reality, a malleable mechanism through which extremist groups seek to draw support from a diverse array of audiences simultaneously rather than an ideological monolith through which they signal costly devotion to a small set of causes, then our current policy responses to extremists targeting only ideological or brutally violent content are likely to be insufficient in addressing extremist appeal, popularity, and enduring influence. Understanding extremists as primarily *political*, rather than *ideological*, actors, can help inform not only our academic understanding of these groups and their enduring influence across the globe, but also help us better adapt and respond to their shifting messaging strategies in conflicts fought nearly as much online as on the battlefield in an increasingly information-centric age of warfare.

2. Measuring Messaging Strategies in Extremist Media: A Content Analysis of Islamic State Propaganda

2.1 Abstract

The proliferation of online propaganda has become a dominant feature of extremism in the 21st century – dominating not only major headlines in mainstream media coverage, but also serving as a highly successful marketing and recruitment strategy for fledgling and experienced groups alike seeking to expand their external audiences beyond their areas of operation. In this paper, I analyze this strategy through an emerging methodological framework for analyzing online content produced by extremist groups online that employs both unsupervised and supervised machine learning methods to the analysis of large volumes text, audio, and video data. I explore the potential underlying content dimensions of Islamic State propaganda across propaganda mediums while illustrating a key application of the more theoretically informed supervised methods to classify, label, or score this data as a series in accordance to the pre-defined characteristics associated with each category. I apply this framework to an analysis of Islamic State propaganda across eleven countries from 2014 to 2018 – the most comprehensive analysis of Islamic State propaganda using these methods to date.

2.2 Introduction

Extremist media is often talked about in either frequency or in anecdote, but we lack, to date, a comprehensive understanding of how the content of any one group or groups itself varies over time and across geographic space. This is problematic for at least two reasons. First, it leaves

us vulnerable to media generalizations about content produced by groups gleaned from a small set of sources. This leads to an understanding of extremist media that remains at the whim of the popular *narrative du jour*, oscillating from violent to repentant to brooding at the mercy of the news cycle. Second, it poses a major roadblock to both theory building and analysis for social scientists looking to make sense of patterns in media. Without a theoretically informed and empirically rigorous way of identifying patterns in high-frequency media data, social scientists may be left with no place to start and no rigorous empirical direction to take in analyzing group behavior in the domain of media strategy.

At the heart of these problems is a critical threat to our understanding of extremist behavior that generalizes to many other domains sharing similar data characteristics: without a more systematic picture of the data available and a baseline understanding of variation, we lean on assumptions built on incomplete data – limiting the issues we can credibly speak to despite an enormous frequency of observations and anecdotal accounts pointing in several promising directions. In the case of the Islamic State, we are fortunate to have several credible "leads" informing our understanding of their media strategy playbook from pioneering scholars delving into their volumes of propaganda by hand. What we lack, however, is a way to construct, organize, and make meaning of high-frequency media data encompassing video, audio, and text releases across dozens of languages and distributed worldwide on the internet.

In this paper, I employ a text-as-data framework to solve this problem through the application of machine learning methods to explore substantive variation in the video, text, and audio data comprising extremist media. Using "bag of words" techniques to reduce content into analyzable form, I first lay out and demonstrate the use of unsupervised techniques for a richer discovery of the potential axes of variation to be found in a pool of data, then dig deeper into

theoretical foundations before employing supervised techniques to classify extremist media content – and, most specifically, media data produced by the Islamic State – across theoretically informed content categories corresponding to my formulation of scholars' best understanding of how extremists use media to attract followers and detract resistors to their causes.

This paper is significant for a variety of reasons. First, it presents and demonstrates the benefits of an applied framework of discovery in text analysis in the social sciences – a framework derived from the pioneering work of Grimmer, Roberts, and Stewart (2022) and extending deep into the essays that follow in this dissertation. Second, it applies this framework to high frequency data on the most relevant extremist group of our lifetimes to date: the Islamic State. In doing so, it represents the most comprehensive machine-learning-driven text analysis of Islamic State propaganda on record so far. While standing on the shoulder of giants, this work and the succeeding two papers pave the way for future advances by significantly improving our understanding of how to learn about extremists through their media strategies.

2.3 Applying Machine Learning Tools to Extremist Media

Machine learning tools have greatly influenced political science research over the past two decades. These methods gained traction not only for their ability to eliminate the resource and time costs of manually sorting through data, but also because of the flexibility of different tools available to scholars who employ them. From data collection, to discovery, to analysis, to robustness checks, machine learning tools in the form of intelligent search automation techniques, classifiers, PCA, and neural networks have become invaluable parts of the research arsenal of scholars discipline wide.

Perhaps nowhere have these tools had more of an impact than in the analysis of text data – a critical feature in my own research. Early innovation in text analysis in political science used these tools to analyze high-frequency text data through naïve classification algorithms to answer decades-old questions of critical document authorship (Holmes & Forsyth 1995), textual interpretation (Bruninghaus & Ashley, 1997), and the classification of media articles into categories using unsupervised techniques (Crawford et al., 1991). More recent advances by pathbreaking scholars have unlocked the potential of machine learning applications of text analysis in the study of high-frequency social media data (King et al., 2013, 2017), the study of legislative claims vs. behavior (Grimmer. 2012), the behavior of jurists in Chinese courts (Stern et al. forthcoming), mass state-sponsored internet censorship in China (Roberts, 2018), the reactionary radicalization of immigrant communities in Europe (Mitts,2019), and many, many more.

Hidden beneath the umbrella of machine learning based text analysis methodology is a diverse toolset to be applied to different stages of the research process. Chief among these tools are three different types of learning frameworks: unsupervised, supervised, and reinforcement learning. Unsupervised learning seeks to glean underlying relationships between unlabeled data points – that is, data points without a coding or designation for a key value or values of interest through a variety of means, such as clustering, component analyses, factorization, or value decomposition (Hinton & Terrence. 1999; Hastie et al., 2009). Supervised learning, by contrast, seeks to train an algorithm to accurately derive the relationships between a set of potentially important features and a labeled value for a feature of interest for the purpose of accurately predicting the value or category of that feature in unlabeled data (Russel & Norvig, 2010).

In this dissertation, I adapt tools from this toolbox to the complex task of empirically analyzing the distribution of substantive appeals in extremist propaganda – most specifically,

propaganda produced by the Islamic State across three continents from 2011-2021. I employ two tools from this toolbox at two different stages of the research process elaborated here. The first tool is discovery – the application of unsupervised machine learning to discover and organize a sample of high-frequency data around latent content dimensions that provide me with a broader understanding of the potential axes of variation, distribution, and importance in my data (Grimmer et al., 2021). The second tool is machine learning- empowered operationalization of my data through the application of supervised learning to classify the releases in my corpus into theoretically-determined categories based on latent features of the content.

This framework pairs a carefully structured discovery period allowing researchers to learn from the data with a theoretically grounded understanding of how extremists may seek to connect with diverse audiences across populations they seek to influence. Furthermore, it marries such discovery and theoretical foundations with methods designed to leverage theoretical insights to classify thousands – and potentially millions – of texts, audio files, and even videos into theoretically informed content categories.

2.4 Application: Media in the Islamic State

To characterize rebel media strategy, I draw on media data predominantly collected from scholar of jihadist movements Aaron Zelin's primary source clearinghouse jihadology.net, which seeks to collect primary source propaganda materials from Salafi-Jihadist groups online and is used in several cutting-edge empirical studies of jihadist propaganda (Zelin, 2015; Walter & Phillips, 2019; Mitts, Phillips, & Walter, 2022). I scrape this source for all media produced between 2011, the earliest date of release available for the Islamic State, the primary subject of my within-case analysis, and 2018, the point at which most key organizations substantially slow the

distribution of propaganda through online channels due to increased online censorship on social media – pushing much of their distribution of media content from public channels to private channels that limit the empirical scope of inference (Twitter, 2018; Mitts, 2019).

I focus the data collection efforts and quantitative analyses on media produced by the Islamic State between 2014 and 2018. I rely on *this* data from *this* range of time for three reasons. First, the richness of this data – over 5,600 independent video, audio, and periodical releases distributed and available through publicly accessible channels over five years – allows me to conduct a fine-grained analysis of media behavior not possible with other groups. Second, the decentralized structure of the Islamic State's media bureaucracy allows me to identify the region of origin of media releases with a higher degree of certainty than possible with other groups in my data, which in turn allows me to explore hypotheses concerning the geography of media production and distribution. Third, this range of time allows me to focus my analysis on the *public* dissemination of rebel media because prior to late 2017, groups like the Islamic State were still able to disseminate propaganda content easily on social media hubs like Facebook and Twitter without widespread pushback against their methods of dissemination from the companies themselves (Mitts, 2019).

The unit of analysis in this case – the parameter that describes what each observation in my dataset used for analysis represents – is the individual media release. This data contains three different types of releases: video, audio, and text file releases. Table TK contains descriptive information about the breakdown of data included in this dataset, while Figure TK displays examples of each of these three types of media.

Table 1: Descriptive statistics of types of media in Islamic State media dataset used in this analysis.

Media Type	No. of Releases	Prop. of Releases
Video	980	0.76
Audio	105	0.08
Text	208	0.16

2.5 Preparing Islamic State Media for Content Analysis

Detecting thematic content in a corpus of thousands of text, audio, and video releases poses a challenge to researchers. Relying on human coders to read, listen to, or watch each release to translate or transliterate content – or even to categorize the content into categories for supervised learning – would require potentially hundreds of labor hours, and in the context of rebel media, may expose human coders to potentially traumatic content. In order to identify thematic appeals within rebel media without relying exclusively on human coders, I turn to computational tools – IBM's Watson transliteration service and Amazon's Rekognition object identification API – that allow me to strip each type of release into a bag of words and images that allows coders and machine algorithms alike to classify the thematic content of each release without directly observing the media (Mitts et al., 2021). Watson transliterations turn every word spoken in a video into text. Rekognition identifies each object within a $1/10^{th}$ second frame of a video.

Raw Release

IBM Watson Transliteration

وهي بازدياد وانظمة اه صحيح ابو رائد مهما كان بس ليلة عنيز بيل سبيكس رجب
بلديات استريبو مشترك تميزت بريسينو لما غيرنا استاذ بريش ومن روبروا يوصل
كيكا فريخ حتى استريز امامنا كافكا زيباكويرا اشتركت فيها لأن بودوي والكرواتي
افترار يستثني بيروت تجري وشن يجد كان ردود وغيرها ونستورد بيصير غير...

Amazon Rekognition Output

Aircraft	Bomb	Chain Saw	Fire Screen	Kicking
Ammunition	Bomber	Cuff	Fist	Knife
Archery	Bonfire	Cutlery	Flame	Lava
Armor	Bow	Dagger	Flare	Machine Gun

Figure 1: Three sources of information available to coders when deciding which content category to code a particular video release. Videos comprise the majority of samples in the content corpus.

2.6 Unsupervised Learning

2.6.1 Model selection

Unsupervised learning methods allow scholars to gain insight into features of their data that may inform theory building, future data collection and refinement strategies, and future empirical designs. Several existing methods allow researchers to conduct any number of research tasks – from revealing previously undiscovered content categories in their data (Grimmer et al., 2012), to reducing dimensionality to improve inference in statistical analysis (Ahlquist & Breunig, 2012; Blaydes & Linzer, 2008), to admixture models that seek to discover clusters of related material in latent content categories in (predominantly text) materials (Grimmer et al., 2022).

I employ an admixture approach using a structural topical model to uncover clusters of related content along latent content dimensions in my text (video, audio, and written media) corpus of Islamic State propaganda in this analysis. Structural topical models generalize several popular features of unsupervised models (Grimmer et al., 2012) and provide several intuitive and interpretable outputs that allow scholars to derive more flexible insights from their sampled data. The STM package provided by Roberts et al., (2016) also allows for basic descriptive and regression features for identifying relationships between characteristics of your data and latent content categories that I employ in my analysis.

Structural topical models require researchers to take several steps before analysis. The first steps involve light-to-extensive pre-processing – creating a text corpus in which the individual texts are stripped of analytically-less-informative features like punctuation, stopwords (the, and, etc.), and tense characteristics. Preprocessing is a non-trivial process – I employ the removal of punctuation, stop-words, and tense characteristics in this analysis.

Researchers can then take several steps beyond this process to further refine the process. In particular, for Arabic language content in the supervised learning section that follows, I applied the package `arabicStemR` (Nielsen 2017) to conduct analysis on Arabic language text that conformed to the structure of the Arabic language in order to preserve useful substantive meaning during the preprocessing step. Because the labeling in the supervised learning section was conducted in Arabic, this was done to ensure that the model captured similar associations between words and content as the human coders did.

2.6.2 Feature covariates

One of the most critical steps in training a structural topical model is the selection of features that might affect what topic each text may belong to. In this model, I allow topical clustering and prevalence to vary according to three covariates. We are, in effect, identifying features that may be related to the clustering of our topics, and the prevalence of those topics in the corpus. I identify the following parameters:

- Time, represented as the date of release for each media observation. Time is critical because we may expect that extremists would release propaganda of particular topical themes at similar times – suggesting that we should try to capture topics clustered in time in our analysis.
- Medium of content, represented as a categorical variable representing video, audio, or text releases. Medium is also an important element that may determine content categories, since certain types of content – religious chants, radio battle updates, or dynamic scenes of battle, for example – may be more likely to cluster along the same topical categories. Note that this variable is evaluated as three dichotomous variables (video, audio, and text) in the regression analysis.
- Governate-level attribution, represented as a dummy variable indicating that the release is attributed to a particular governate-level location in the Islamic State. The governate in which the propaganda is produced, and from which it is disseminated, may focus specifically on certain types of appeals that the group seeks to target at particular audiences to get them to associate the group's presence in that territory with a particular topical theme. These appeals may generalize to more "local", identifiable content categories, and should be accounted for.

2.6.3 Determining number of content categories

A final critical step in training an STM is determining the number of topical clusters to identify in the discovery process. To a certain extent, this process is fundamentally arbitrary; in a true discovery process, using this method to uncover unknown variation in the topics in your data implies a particular lack of knowledge about the topical dimensions of the data with which you are working. In order to ensure that I provided the model with enough categories to capture variation in over 5,000 releases while consolidating enough to prevent topical overlap, I set topics (k) to 10 for this analysis.

2.6.4 Results and Interpretation

Table TK shows the labels corresponding to the ten categories produced by the STM analysis, provides an interpretation of what each of the identified ten topics symbolize, and displays the proportion of each topic in the corpus. An important feature of structural topical modeling is the reliance on human-selected labels reflecting the researcher's approximation of what the words grouped together by the clustering topical model express thematically. In order to provide more insight into how I selected each topic label, I will briefly cover each topic category, justify my naming of this topic, and provide an example of media falling into this topic category based on this scoring:

1. Proclamations of legitimacy: This topic label captures the high prevalence of language around the recognition of the Islamic State as a religious and political body under the notion of a Caliphate, or *khilafa* – and the then-leader of the Islamic State Abu Bakr al-Baghdadi as the legitimate Caliph, or *khalif.* High-leverage words such as allegiance (*bay'a*), *khalif*, *khilafa*, recognition, legitimacy indicate the importance of allegiance and the authoritative

role of the Caliphate structure as key tenets of legitimacy being drawn upon by media espousing this content. An example of a media release with a high proportion of content corresponding to this category, "On the Allegiance of the Tribes of Irbil to the True Khalif", features several local tribal leaders pledging their allegiance to al-Baghdadi before cutting to a sermon by al-Baghdadi discussing the "virtuous revival" of a global Caliphate led by the Islamic State.

2. Calls to Arms: this topic label reflects the high prevalence of language in releases calling for individuals to take up arms to support the Islamic State and it's causes. High-leverage words such as *fighter*, *homeland*, *support*, *liberator*, and *resistance* indicate the importance of rallying cries to local individuals to take up arms for the Islamic State's cause. An example of a media release with a high proportion of content in this category, "Virtuous Soldier Calls for Righteous Recruits to Join the Just Cause," prominently showcases this topic through appeals by a soldier spotlighting the need for recruits to join him in fighting for the Islamic State.
3. Operational Updates: Contrasting with the Calls to Arms topic, this topic label predominantly captures language related to reporting on the number and types of military operations in which Islamic State militants have participated. High leverage words such as *operation, report, strikes, tally,* and *statistics* display the types of language that we would expect in a summary or dashboard report of operations conducted by the Islamic State and reported through media.
4. Appeal to Foreign Soldiers: in contrast to the content in the Calls to Arms category, appeals to foreign soldiers represents content that specifically appeals to individuals outside of territories with Islamic State military operations and consists of religious and social appeals

for them to travel to participate in Islamic State military operations. High-leverage words such as *overseas, travel, pilgrimage, obligation,* and several country and regional names likely reflecting places of origin for these calls indicate an appeal to the religious and social obligations of individuals from abroad to join the Islamic State's cause. An example of a release with a high proportion of content in this category, "Call to the Righteous Abroad to Join Their Brothers on the Front," features testimony from a group of Islamic State soldiers urging their "brethren" to join them in fighting for the Islamic State in Syria.

5. Religious Violence: this topic label captures religious justifications of violence perpetrated by Islamic State or Islamic State-aligned militants. High-leverage words such as *martyr, glory, jihad, retribution, punishment, non-believer,* and *heavenly reward* reflect the emphasis on religious justifications for violence in the name of divine retribution or punishment – whether the focus of that violence be on the righteous perpetrator or the unrighteous individual punished. An example of a media release with a high proportion of content in this category, "Righteous Punishment for the Infidel (non-believer) in Irbil," displays the features of this category through a description (and depiction) of the execution of an individual for the crime of blasphemy in territory under Islamic State control.
6. State Affairs: This topic encapsulates the administrative, bureaucratic, and "governance" services provided by the Islamic State in areas under their administration, contestation, and control. High-leverage words such as *administration, govern, office, policy, legislate* exemplify the emulation of the state that the Islamic State is seeking to advertise from areas under its control. The media release "Office Regulating the Administration of Taxes in the Caliphate in Raqqa" provides an example through the portrayal of governance institutions in a state capacity.

7. Programmatic Updates: This topic reflects social, economic, and religious services and governance programs/initiatives launched by the Islamic State. High-leverage terms such as *zakat* (Islamic charity regulated by the state), *regulation, program, inspection,* and *traffic* reflect the programmatic center-of-gravity of this topic. The release "Inspection of Meat Quality by Officials in Accordance to Righteous Law" features an Islamic State appointed inspector commenting on the regulation of meat production and sale in territories under Islamic State control alongside featured testimony of residents testifying to the claimed improvement in meat quality since such inspections.
8. Provincial Spotlights: This topic reflects rebuilding, new construction, and returning critical infrastructure projects in provinces under Islamic State control. High-leverage terms such as *reconstruction, project, buildings, new (building), renewal, rebuilding* and *province* spotlight the emphasis on rebuilding areas that were affected by conflict. The media release "Rebuilding Houses Destroyed by the Infidels in Raqqa Province" displays this emphasis on highlighting Islamic State reconstruction and renewal.
9. Threats to West: This topic showcases threats of violence, retribution, revenge, and future harm toward politicians, soldiers, and civilians in the West – a broad grouping typically singling out America and Western Europe, but occasionally including Russia, China, Japan, and Australia. High-leverage terms such as *revenge, retaliation, smite, hammer, foreign* and *West* capture the spirit if this topic. An example of a media release with a high proportion of content in this category, "A Promise of Revenge to Civilians in the West Supporting Oppression of the Righteous," depicts a nearly-thirty-minute list of threats of violence to civilians in the West who support politicians that take military action in the

Islamic State before playing footage of the Bataclan Theatre terrorist attack that occurred in Paris, France in 2016.

10. Enemy Atrocities: This topic brings attention to atrocities committed by Western or pro-Western government forces against Sunni Muslims. The high-leverage terms *massacre*, *atrocity*, *fratricide*, *murder*, and *betrayal* indicate the prevalence of atrocious violence – sometimes committed by individuals from the same religious group ("fratricide" is a commonly used term in religious rhetoric to describe the killing of Sunni Muslims by other Sunni Muslims). A media release containing a high proportion of content from this category, "Massacre of Our Brothers by the Nonbelievers in Shmaal Baghdad," features the depiction of dead soldiers after an apparent air strike in territory contested by the Islamic State in the region north of Baghdad in Iraq.

Some of the insights that this analysis reveals are unsurprising: topics organized around calls to arms, appeals to foreign soldiers, and updates on battlefield operations account for over one third of the video, audio, and textual content captured in this analysis. We might expect an organization recruiting actively from overseas and continuously mobilizing its forces for battle to create such content in large proportions. However, a number of surprising features compliment these findings. First, content sharing features associated with the "religious violence" label – often the content most highlighted in media portrayals of the Islamic State – accounts for a surprisingly small proportion of overall content at 9%. In fact, topics coalescing around violence overall seem to account for a very small minority of topics discovered by this model – accounting for just 18% of the corpus in this analysis. As I explore in the second and third papers of this dissertation, the relationship between media and violence in Islamic State media strategy is far more complex than

it is often portrayed. By using this method, we can uncover surprising variation like this and unlock topics of future exploration using more theoretically grounded, deductive methods of research.

Second, this analysis suggests that the Islamic State devotes a substantial amount of content to updating followers on its programmatic, bureaucratic, and service provision functions: the topics capturing state affairs, programmatic updates, and provincial spotlights account for a substantial 23% of the content in this corpus. This focus on content related to governance and territorial upkeep points in the direction of a little understood phenomenon: how, when, and why do extremists advertise and portray governance? I tackle this question in the second paper of this dissertation by breaking down the operational – and territorial – incentives to shift media strategies among extremist groups. Discovery has, yet again, revealed a potential latent dimension in this content worthy of future deductive exploration.

Finally, the most prevalent topic in this analysis – a topic I have labeled "proclamations of legitimacy" – organizes language indicating the agglomeration of social, political, ideological, religious and economic praise for Islamic State success. The combination of ideologically charged language along dimensions of social, political, and economic success indicates the potential for indoctrination around a shared set of principles through highlighting the success of the Islamic State's model. I explore this potential further in the second paper of the dissertation as well.

Table 2: Topics from the STM model of topics in the Islamic State corpus. This table provides the rank of each topic, the label we have provided for it based on high-prevalence and high-leverage words occurring in this topic, an interpretation of each label, and the proportion of each topic in the corpus.

Rank No.	Topic Label	Interpretation	Prop.
1	Proclamations of Legitimacy	Social, political, religious, and economic praise for the emerging Islamic State as a governing body	0.16
2	Calls to Arms	Calls to local fighters to defeat the enemy and rally around the Islamic State's cause	0.15
3	Operational Updates	Updates on military operations across areas of combat engagement and rebuilding projects	0.13
4	Appeal to Foreign Soldiers	Religious and social appeals to foreign combatants to join the Islamic State (in their territory)	0.12
5	Religious Violence	Religious glorification of combat violence	0.09
6	State Affairs	Announcement of expansion, new administrative function, or new military, social, or economic policy	0.08
7	Programmatic Updates	Updates on ongoing social, economic, and charity programs led by the Islamic State in its territories	0.07
8	Provincial Spotlights	Content spotlighting critical rebuilding or service providing projects in Islamic State territories	0.07
9	Threats to West	Threats of retribution, violence, and revenge on Western forces and civilians	0.05
10	Enemy Atrocities	Spotlights of atrocities committed by enemies against local Sunni Muslim communities	0.04

In addition to discovering unexpected dimensions of variation in the topics output by this analysis, the results on the covariates included in the model also introduce intriguing variation that raises questions for future analysis. In particular, the medium of release and governate-produced content variables prompt further discussion for further exploration. Of the ten content categories identified, four (Calls to Arms, Provincial Spotlights, Threats to the West, and Enemy Atrocities) were significantly more likely to be released on video, while two (Religious Violence and Proclamations of Legitimacy) were significantly more likely to be released over text format. Likewise, State Affairs, Programmatic Updates, and Provincial Spotlights were widely and significantly more likely to occur at the governate level – a finding that implies more local ownership and marketing of governance-oriented releases.

These findings raise interesting questions about the role of different types of media in accomplishing the strategic messaging goals that extremists may have – a question for future research to explore. In addition, these results suggest that governate-level releases are more likely to feature governance-adjacent topics such as state affairs, programmatic updates and – least surprisingly – provincial spotlights. This finding reveals the potential for a connection between media strategy and territorial control – another example of the discovery process leading to nuggets of future research through other means. I explore this phenomenon in greater detail in the supervised learning section below and paper two of this dissertation.

2.7 Supervised Learning

Supervised learning differs from unsupervised learning by inserting a human into the loop to synthesize theoretically meaningful categorizations of data. In supervised learning, a human coder will take a theoretically motivated set of categories and label data across these categories – deciding which content belongs in which categorical bucket, or otherwise assigning metrics across theoretically informed categories. These labels are then used by a learning algorithm, along with the data, to train a model that then classifies content based on the underlying relationship between the content inputs and the labelled associations affixed by the human coders. In the section that follows, I break down this process, beginning with the content categories selected and moving through the data splitting, human labelling, model training, classification, and performance assessment processes.

2.7.1 Determining Content Categories

Supervised learning (done well) relies on a theoretically grounded process consisting of identifying the salient content categories based on theoretical expectation of how extremists should

attract followers. This process isn't entirely divorced from the data. We often start with an unstructured picture of data, anecdotal examples of appeals, and information on how groups operate. The important step present in the supervised process from the outset is the human creation and assignment of content features assigned to data. In the case of classification problems such as identifying the thematic categories of propaganda data, the derivation of content categories is a critical part of this process.

The content categories used in this paper represent three theoretically driven categories introduced in the introduction of this dissertation: indoctrination, co-optation, and intimidation. To recap these three categories:

1. Indoctrination: shift or exploit people's existing beliefs to make them subscribe to the same goals, thus motivating them to take up their cause;
2. Co-optation: entice people with promises of physical, political, or social rewards without shifting their ideological beliefs to conform to their goals;
3. Intimidation: force people into compliance with threats of physical or psychological punishment for interfering with or resisting their operations.

I explore the theoretical foundations of each of these three macro-categories in the next essay in this dissertation (see pages 47-55 for a full theoretical exploration). The root of each of these theoretical foundations are what I argue are the foundational strategic tools of persuasion – the tools that extremist groups use to entice individuals to join their cause through either the promise of material, social or religious goods and the promise of a life within a righteous community of likeminded individuals, or else the specter of violence if they do not acquiesce to extremist demands. Thus, each of these categories is borne from a theoretically derived

understanding of the persuasive strategies extremists adopt to communicate with a diverse group of followers who, by virtue of location or circumstance, might be better reached through media communication.

For a further explanation of how coders were trained to recognize these appeals, see Appendix A.

2.7.2 Splitting Data for Learning

Supervised learning requires splitting the available data into different sets to be used at different stages of the learning process. I split my data into a labeled training set comprising of 65% of the available data and a test set comprising the remaining 35%. This labeled training set is used to train the model used to estimate the parameters of content-based model that determines the most likely dominant thematic appeal in each media observation. The test set, in turn, is used to validate the model performance in data withheld from the training of the model to determine the accuracy of a model generated from the training set.

2.7.3 Human labeling

In order to label categories, I instructed human coders (research assistants fluent in Arabic and emerging subject matter experts in the Islamic State) to observe the title, Watson transliteration, and the most prevalent unique objects in a sample of 2,500 releases and observe the accompanying title banner (example above under “Raw Release”) before determining the

significant themes present based on a list of identifiable characteristics.[1] This adapted list of identifiable characteristics separates content into four categories based on each feature's theoretical closeness to the theme described in the section above: indoctrination, co-optation, intimidation, and an "other" category representing content that does not fit into these thematic categories. A distribution of labels can be found in Appendix A.

2.7.4 Model Selection & Training

Next, I train two separate machine learning algorithms– a naïve Bayes classifier and a random forest classifier– on a labeled training set before 1) validating the models on a test set of data, then 2) using these models to classify the data not coded by human coders. The naïve Bayes classifier is a simple Bayesian classifier often used as a baseline in classification tasks because of its generalizability, relative ease of interpretation, and widespread use in the machine learning literature. Likewise, the random forest model is one of the most frequently used classifiers in the applied machine learning literature, makes well-understood assumptions, and can generally achieve high out-of-sample accuracy, precision, and recall. For both models, I use 75% of the available media data to train a model as a training set, setting aside the remaining 25% of my data for validation of out-of-sample performance.

[1] Thank you to Haifa Albeleihid, Mihiri Kotikawatta, and Thomas Nemechek for their outstanding research assistance.

Table 3: Out of Sample Naïve Bayes Classifier Performance

Category	Precision	Recall	Accuracy (F1)
Indoctrination	0.89	0.78	0.85
Co-optation	0.84	0.81	0.82
Intimidation	0.79	0.71	0.73

Table 4: Out of Sample Random Forest Classifier Performance

Category	Precision	Recall	Accuracy (F1)
Indoctrination	0.88	0.93	0.89
Co-optation	0.82	0.85	0.84
Intimidation	0.77	0.94	0.84

2.7.5 Supervised Learning Classification Results

The accuracy, precision, and recall of each classifier are provided in Tables 4 and 5. As shown in these tables, both models achieve reasonably high measures of out-of-sample precision, recall, and accuracy (as measured by the F1 statistic reported). In both models, the indoctrination category is classified with the highest degree of accuracy, while intimidation is classified with the lowest degree of accuracy.

It is important to note what these results can and, equally as importantly, *cannot* tell us about how well these theoretically informed categories describe this data. On the one hand, these results indicate that these theoretically informed content categories can be detected in the features

of the media analyzed in the model (bags of words, in this case) with a degree of accuracy much higher than random chance when trained on human labelling. This provides us with a limited degree of evidence that we have identified distinct, identifiable categories of content in line with these theoretically informed features.

On the other, we must content with the subjective reality and empirical limitations of such classification techniques. While this high degree of accuracy in out-of-sample accuracy provides us evidence that these theoretically informed categories are distinguishable from one another (and any other latent categories present in the data), it does not provide any degree of prima facie evidence that these categories 1) mean what we think they mean, or 2) reflect the underlying intent of the content creators when generating the content. The beauty and the burden of these methods is their flexibility and, ultimately, their ambiguity, and the burden rests on scholars to justify their selection of classification categories and their use in empirical analyses. I engage with the theoretical importance of these categories in the discussion below, and in far greater detail in the next paper in this dissertation.

2.7.6 Analyzing These Results in Context

The results reported in table TK illustrate the performance of these classification models in identifying salient themes in Islamic State media, but what can these results tell us about the group's strategic use of this media? In this section, I briefly introduce a descriptive illustration of the usefulness of this data: a visual chronology of Islamic State media themes and a provincial breakdown of thematic distributions across Islamic State provinces under their contestation or control between 2014 and 2020. These analytical applications illustrate the types of questions that text analysis helps us to uncover and motivate.

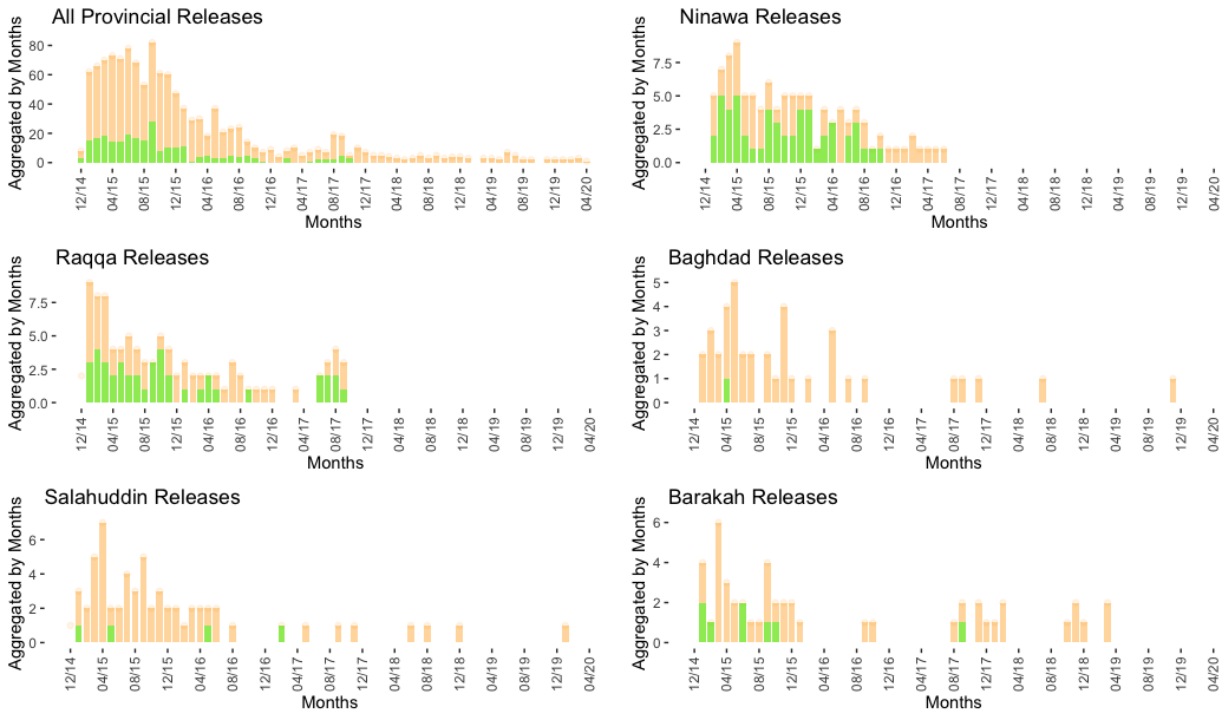


Figure 2: Proportion of all local-level co-optation (A), indoctrination (B), and civilian intimidation (C) releases produced by the Islamic State across all governates (top left) and five selected governates. The Islamic State controlled significant territory in two of these governates (Ninawa and Raqqa), never took over significant territorial holdings in two others (Baghdad and Salahuddin), and established a very limited and short term presence in Barakah governate.

Figure 2 shows a distribution of the three thematic categories representing messaging strategies – co-optation, indoctrination, and intimidation – from 2014-2020 across different provinces across Iraq, Syria, and Libya either contested by or under the control of the Islamic State during this period. This figure highlights four distinct sources of variation across our content categories: variation within content categories over time, variation across content categories over time, variation within content categories across geographic space, and variation across content categories across geographic space. These sources of variation may lead us next to insightful research questions about how groups like the Islamic State use media strategically across time and space – and why they might employ more of a certain type of content at one time or in one place, then lean more heavily on another otherwise.

For example, the concentration of co-optation based media produced in (and attributed to) Islamic State operations in Raqqa (from the start of our data collection to late 2017) and Ninawa (from mid-2014 to early 2016), two regions governed by the Islamic State during this period, and the relative dominance of this category in these regions both when compared to other provinces in which the Islamic State had no foothold and compared to other content categories attributed to these areas, suggests a potential relationship between territorial control and media strategy that may merit further theoretical exploration. This relationship between the content categories – both across time and space, and relative to one another – serves as the basis for the theoretical and empirical investigations in the following two papers of this dissertation. But even more than serving as the foundation for the following papers, this applied demonstration reveals the real advantage that tools like supervised learning can provide when applied to data in this way. They allow us to uncover previously hidden layers of variation in the data – illuminating potential paths for further investigation. Used carefully, this tool can empower researchers to better understand which directions to take in research's garden of forking paths.

2.8 Discussion

In this paper, I have argued that applying text-as-data techniques that harness the power of machine learning to efficiently uncover latent dimensions of text and video data can enhance our understanding of extremist behavior – namely, how extremists use media strategically to communicate a variety of messages through their online engagement. Using transcription and transliteration programs to break down thousands of media releases into "bags of words and objects" for analysis, I have demonstrated that propaganda data by groups like the Islamic State provides a workable – and perhaps even ideal – use case for the application of these methods. Through an unsupervised analysis of this media corpus, I have illustrated the usefulness in

techniques like structural topical modeling in allowing researchers to carefully explore their data and potentially discover latent content categories lurking in infeasibly large data lakes of extremist media that may not be efficiently (or ethically) read through and identified by hand. Finally, I contrast this unsupervised approach with a more theoretically-grounded supervised learning approach – inserting human reasoning, theoretical deduction, and labelling into the learning loop – that combines expert recognition of theoretical foundations of extremist behavior with computational efficiency and machine learning capabilities to efficiently and (reasonably) accurately extend our ability to categorize large data sets.

Each of the empirical approaches presents a distinct tradeoff that is worth considering. While breaking down propaganda data into "bags of words and objects" allows us to analyze this data more efficiently with computational resources, it also strips media data of many of the other salient features – tone, aesthetic design choices, motion, and cadence, to name only a few – that we use to make meaning of the content we as humans encounter in the real world. While contemporary scholars are seeking solutions to this problem as we continue to use these tools (for example, see Grimmer et al., 2021 and Lucas & Stewart, 2021), it is very likely uncontroversial to say that this approach necessarily represents a tradeoff between the richness of meaning in content and the efficiency with which we can analyze it. As discussed above, unsupervised learning techniques allow us to recover an enormous amount of meaning through the underlying associations between the features identified in content, but this comes with significant risk of injecting meaning into random, arbitrary, or convenient results if we don't acknowledge the subjectivity of any given chosen hyperparameters or biases of individuals making meaning of the results. Supervised learning techniques allow researchers to radically increase the efficiency of

classification tasks, but they rely on the theoretical understanding and careful, ultimately subjective discernment of human researchers to inform the basis of content being classified.

Ultimately, this paper takes these tools at face value – acknowledging the risks that each pose to inference and carefully engaging in discovery that helps motivate the research that follows in the remaining two papers of this dissertation. From the structural topical model employed in the unsupervised analysis, we learn that the content corpus produced by the Islamic State examined here contains a veritable cornucopia of content themes – many of which appear to have little to do with the type of religious violence associated with the group. From the models employed for our supervised analysis, we take this discovery onto the path of inference by further by motivating our search for meaningful variation in our data through a theoretically informed labeling and classification of the media dataset. This analysis further illustrates the diversity of content, and allows us to analyze the variation across time and space in a more theoretically grounded way – pushing us to explore deeper questions linked to our best understandings of how extremists use media as a strategic tool. In the next paper, these content categories and the variation observed in our applied example above serve as the basis for a deeper investigation into how the Islamic State – and perhaps even extremists at large – use media strategically across space and time.

2.9 Chapter Acknowledgements

Chapter 2, Measuring Message Strategies in Extremist Media, is a solo authored paper. This paper has its origins in my final paper assignment for Molly Roberts' Text As Data course, and I humbly acknowledge Prof. Barbara Walter's, Prof. Dotan Haim's, and Prof. Molly Robert's contributions to this work through their employment, collaboration in very early drafts, and feedback/teaching respectively. For that project, I worked with Dotan Haim to conduct simple

unsupervised and unsupervised content classification tasks on a smaller subset (2 years) of extremist propaganda data. We conducted these analyses on data I had collected in 2015-2016 as a research assistant for Prof. Barbara Walter, along with supplemental data collected following the end of that RA work collected by me.

This analysis differed from the paper presented here in three critical ways. First, the scope of the data used in the analysis was far more limited; whereas we used only text data in the form of titles for the analysis in the course paper, I take advantage of both the transcribed words spoken in videos, audio releases and texts, and the identified objects identified in text, in this paper. Second, the sophistication of methods employed is greater: I approached this paper with a far more theoretically grounded set of categories corresponding to the theoretical heart of this dissertation (Indoctrination, Co-optation, and Intimidation) rather than the far earlier appeals brainstormed for the term paper, and used an expanded scope of machine learning methods for my supervised learning analysis. Third, this paper was written entirely on my own, without contributions by any of the above in the writing process.

2.10 References

Ahlquist, J., & Breunig, C. (2012). Model-based Clustering and Typologies in the Social Sciences. *Political Analysis, 20*(1), 92-112. https://doi.org/10.1093/pan/mpr039

Grimmer, J., Messing, S., & Westwood, S. (2012). How Words and Money Cultivate a Personal Vote: The Effect of Legislator Credit Claiming on Constituent Credit Allocation. *American Political Science Review*, *106*(4), 703-719. https://doi.org/10.1017/s0003055412000457

Grimmer, J., Roberts, M., & Stewart, B. (2022). *Text as Data: A New Framework for Machine Learning and the Social Sciences* (1st ed.). Princeton University Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (1st ed., p. 764).

King, G., Pan, J., & Roberts, M. (2013). How Censorship in China Allows Government Criticism but Silences Collective Expression. *American Political Science Review, 107*(2), 326-343. https://doi.org/10.1017/s0003055413000014

King, G., Pan, J., & Roberts, M. (2017). How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument. *American Political Science Review*, *111*(3), 484-501. https://doi.org/10.1017/s0003055417000144

Lucas, C., Nielsen, R., Roberts, M., Stewart, B., Storer, A., & Tingley, D. (2015). Computer-Assisted Text Analysis for Comparative Politics. *Political Analysis*, *23*(2), 254-277. https://doi.org/10.1093/pan/mpu019

Mitts, T. (2018). From Isolation to Radicalization: Anti-Muslim Hostility and Support for ISIS in the West. *American Political Science Review, 113*(1), 173-194. https://doi.org/10.1017/s0003055418000618

Mitts, T., Gregoire, P., & Barbara, F. Extremist Radicalization and Recruitment in an Internet Age. *Revise And Resubmit At Journal Of Politics*, (Working Paper).

Nielsen, R. (2017). *ArabicStemR Package* [Ebook] (pp. 1-19). Retrieved 21 March 2022, from https://cran.r-project.org/web/packages/arabicStemR/arabicStemR.pdf.

Russell, S., & Norvig, P. (2009). *Artificial intelligence* (3rd ed.). Pearson.

Walter, B. (2017). The Extremist's Advantage in Civil Wars. *International Security*, *42*(2), 7-39. https://doi.org/10.1162/isec_a_00292

Walter, B. F., and Gregoire, P. (2018). "Rebel Online Propaganda in an Internet Age." *Under Review*.

3. Mainstreaming Extremism: Indoctrination, Co-optation, and Intimidation on the Battlefield of Hearts and Minds

Gregoire Phillips

Department of Political Science

University of California, San Diego

3.1 Abstract

While much of the contemporary coverage of extremist media produced by groups such as the Islamic State focuses on displays of violence and brutality, a more comprehensive thematic analysis of the media produced by such organizations reveals a surprising amount of variation in the content produced. In fact, the themes employed by the Islamic State varies substantially not just across time, but also across the regions from which they attribute their media production. How can we explain the substantial variation in content produced by extremists across time and space, and how do messaging strategies shift in response to operational realities for such groups on the ground? In this paper, I propose a novel strategic framework for understanding how extremists shape their messaging strategies built around three foundational strategic frames: indoctrination, co-optation, and intimidation. I then evaluate this framework on an empirical analysis of the effect of territorial control on messaging strategies employed by the Islamic State in Iraq, Syria, and North Africa from 2014 to 2018. I find that territorial acquisition significantly shifts messaging strategies from indoctrination and intimidation toward co-optation – reflecting a mainstreaming

shift toward material and governance content as extremists shift from conquerors to stewards of lands and communities. I conclude with a brief discussion of these results and phenomena.

3.2 Introduction

Much of the existing literature in political science tells us that extremist are unique among rebel groups, and that holding extremist beliefs confers upon them specific advantages (Walter, 2017; Berman, 2011; Sanin & Wood, 2014; Weinstein, 2006) and disadvantages (Shapiro, 2015; Fortna, 2015). However, while extremists are typically very much in a league of their own ideologically, they reside in the same neighborhood as many other violent non-state actors: outside the state, and with explicit political goals to defeat and supplant the government.

Violent extremist organizations are, most typically, a subset of rebel groups. Like all rebel groups, they often have very specific political goals. They need three things: recruits to carry out their operations, financing to pay soldiers and acquire the machinery of war, and information on the capabilities and operations of their opponents (Berman et al., 2018). Unlike all other rebel groups, however, extremist goals – and the means they are willing to use to accomplish them – typically reside far outside the ideological comfort zones of most of the population. Espousing or publicly supporting extremist goals can, almost without question, expose people to violent social or political backlash.

This wouldn't be a problem if extremists could rely only on true believers, or those purely devoted to their cause, to accomplish their goals. If those goals are limited enough – limited to small-scale operations to accomplish very specific political or social change – this might be enough to get the job done. However, for most groups, their goals are more expansive – from reviving a medieval political order to expelling or exterminating an entire group of people from society. This

makes reliance on true believers (Roeder, 2017) – individuals who are already indoctrinated into the cause or causes espoused by extremists – impossible in the long run: there simply aren't enough extremist-aligned recruits or financiers available in most local contexts to keep a successful extremist movement off the ground without expanding beyond the ranks of the first true believers.

Extremists, therefore, have a problem: they need to expand their support beyond the core ranks from which their movements emerge, but doing so involves convincing people to take often-extreme risks to support causes and methods that most of them don't initially agree with. They can't usually campaign in the open from the start of their movement, and in-person communication poses considerable risks both to the group and the potential followers. How can extremists expand their outreach without compromising the safety of their organization or potential recruits?

To tackle this problem, extremists invest in messaging strategies that seek to draw people to their cause without direct engagement. By disseminating information that advertises the groups goals, resources, rewards, power, and punishments for those that stand in their way, extremists can expand their efforts to increase their ranks, fundraise for their causes, and deter opposition without taking the substantial risks that widespread in-person recruitment or open suppression poses.

The content of extremist messaging reflects not just the quality and reputation of the group, but also a snapshot of the strategic logic of the organization at a given point in time. The substance of any given piece reveals an extension of the game extremists seek to play with their target audience: extremists use the content of their media to communicate specific information to their target audience with the hopes that this piece of information helps draw that consumer to the organization or deters them from resistance. Identifying the content and its target audience allows

us to determine what game extremists are playing with which audience – information that reveals much about what extremists need and what they plan to do next on the battlefield.

3.3 What Extremists Need

Extremist messaging games, like all strategic games, can be broken down into four components: actors, interests, strategies, and expected outcomes. Extremists comprise one actor, while the audiences with whom they seek to communicate comprise the other. Understanding the interests of extremists – what extremists need to accomplish their long-term goals – informs a critical component of their messaging strategy.

While extremist goals may vary widely, from ethno-religious sovereignty (Roeder, 2017) to the withdrawal of foreign soldiers from their homeland (Bergen, 2001), to the revival of a medieval Caliphate (al-Baghdadi, 2014), to the realization of racial or national purity, extremist groups generally require four things to thrive: recruits, money, information, and compliance. Attracting people to the cause is critical: without human bodies, extremists lack the person power to carry out their operations – from carrying out clandestine attacks to mounting an effective military effort on the open battlefield. Raising money allows extremists to expand and maintain the flow of human recruits through wages and financial support while also supporting the purchasing power of the organization in the form of a cashflow they can expend to buy weapons or, in limited circumstances, to buy civilian support in areas under their influence or control. Information about their opponents – from where government influence is strongest, to where it is weakest, to which areas are most suitable to hide clandestine operations in unfamiliar territory, to who might be spying on them for the enemy – is a valuable component of extremist movements in areas beyond their local bastions of control. Finally, compliance with extremist doctrine and rule

is critical – not only because of the ideological desires of leadership, but also because of the degree of control that compliance exerts over the behavior of people who *aren't* part of the organization. If extremists can induce people outside of their ranks to comply with their demands by not revealing information about them to the government, or by paying tribute to the group when under their influence or control, extremists can wield authority over even those outside of their immediate grasp.

These four components comprise extremists' core interests: without these things, they cannot effectively carry out their operations and accomplish their goals. These interests are more or less static; extremists never cease needing recruits, for example, nor do they ever stop seeking funding. They set the bounds for extremist demands from their target audiences.

3.4 Target Audiences

Critical to messaging strategy selection for extremists is the audience with whom they believe they are communicating. I identify three audiences that extremists seek to target with messaging strategies: sympathetic, malleable, and oppositional audiences.

A sympathetic audience is one who generally agrees with – or could be easily steered to agree with – extremist goals and the methods they employ to obtain them. In such an audience, we may find the "true believers" committed to a set of goals aligned – or nearly aligned – with the political ideology espoused by the extremist group's leadership. These audiences may be predisposed, by nature of their prior beliefs, to the "platform" of the extremist group by belonging to a particular national, ethnic, or religious group, or by subscribing to a similar worldview (Roeder, 2017). Such individuals may face significant barriers to revealing their support for extremists, such as violent and non-violent sanctioning by their governments or social groups, but

may also be among the most likely to take costly actions for the group (Walter, 2017). These individuals may also be driven into the arms of sympathy for the extremist cause by social isolation or hostility in their home environments (Mitts, 2020).

A malleable audience is one who does not agree with extremist goals, or the methods they employ to obtain them, but can be steered to support extremists anyway in exchange for material, social, or political benefits. These individuals may be driven by material needs, such as poverty, or socio-political discrimination preventing them from equal access to social groups, political institutions, or the marketplace. We may find individuals out of work for months, or potentially years, seeking a change in governing power that will change their fortunes with little regard to the ideology or political cause of such a group. The critical difference between a malleable audience and a sympathetic one is the underlying appeal of the ideological goals of the extremist group to individuals within that audience. A malleable audience isn't driven to support extremists because they believe what that group is doing is *right,* but rather because what that group does could potentially make the lives of people within that audience better relative to the status quo or alternative possibilities under other groups. Support from a malleable audience is instrumental; it relies on the extremist group persuading this audience that they will provide them with goods and services commensurate with the risks of supporting them or acquiescing to their rule.

An oppositional audience is one who does not generally agree with extremist goals or the methods they employ to obtain them, and is additionally unwilling to support the group in exchange for material, political, or social benefits. Individuals in an oppositional audience are unwilling to support extremists out of a commitment to the same or similar ideological principles/goals, and are additionally unwilling to support them instrumentally for material, social, or political purposes. An oppositional argument is driven, in principle, to oppose extremists even

if doing so harms their material, social, or political standing in their communities. Just as a sympathetic audience is potentially willing to forgo benefits to support the extremist cause, an oppositional audience is willing to incur costs to oppose them.

I argue that these three audiences vary on the most critical dimension for extremists crafting a messaging strategy: underlying sympathy for the extremist cause. Conceptualizing extremist audiences in this way allows us to understand how extremists might differentiate the appeal strategies extremists employ in their messaging to gain their support. In what follows, I break down three messaging strategies that respond to such variation in target audiences. I go on to illustrate how these strategies respond to variation in an extremist group's audiences, and how shifts in operational strategy change the audiences to whom extremists appeal – and thus shift their messaging strategies at the same time.

3.5 Indoctrination, Co-optation, and Intimidation

In this dissertation, I argue that extremists can pursue three strategies in pursuit of their interests to recruit followers, raise funds, and deter opposition: indoctrination, co-optation, and intimidation. Each strategy represents a set of actions that extremists pledge to take (or depict themselves taking) alongside a distinct implicit or explicit set of requests or demands that extremists make of their audience. These strategies require persuasive follow-through; selecting a strategy requires extremists to organize their messaging strategies around on-the-ground actions that provide some sort of validity or legitimacy to the claims made on paper, on video, or over the airwaves. As we will see in Chapter 5, the follow through on promises made implicitly or explicitly in media varies drastically across areas controlled and administered by groups such as the Islamic State. Nonetheless, these messaging strategies reflect attempts at persuasion as a strategic response

to 1) the needs of the organization at a given place and time, and 2) the audience from whom they seek resources, recruits, or compliance to fill those needs and accomplish their goals.

3.5.1 Indoctrination

A strategy of indoctrination seeks to shift a target audience's political beliefs to align with an extremist ideology. Indoctrination in this framework refers to the process of teaching people to or demanding that people accept a set of beliefs uncritically. Extremists seek to indoctrinate their audience through persuasion, convincing people that they follow the righteous path to their political destination. Indoctrination, in this sense, requires not just radicalization, or the process of convincing people to unquestionably adopt extremist beliefs about the world, but also the application of those beliefs toward the extremist's political cause.

Scholars of extremism and radicalization have long sought to explain indoctrination as a method of recruitment. L.W. Doob, when laying out his seminal eight principles of propaganda, emphasized the need for a rooting dogma around which to organize people and, through repeated interaction, indoctrinate people to a cause (Doob, 1935). In his later work, Doob describes how Goebbel's, as the Third Reich's principle propagandist, organized the emerging German state information machine around an unquestionable set of ideological principles communicated through repetition, cultural connection, and flattery designed to expand the ranks of true believers in the Nazi cause (Doob, 1950). Funk and Wagnalls (1972) describe indoctrination as a pathway to social control centered around political or sectarian dogmas. Roeder describes the process through which nationalist-secessionist organizations seek to align their political cause for national separation with specific ideological causes, allowing secessionists to indoctrinate "true believers" in that ideological space into their movement (Roeder, 2018).

In the Islamic extremist context, scholars have identified the use of similar tactics of repetition, cultural connection, and flattery in the integration of commonly accepted religious beliefs and common regional identities into radical beliefs about the moral legitimacy of modern states and the justified use of violence against outsiders (Wictorowitz, 2005). Horgan (2008) notes that many of these routes to radicalization succeed because they connect psychological traumas of individuals exposed to violence or discrimination to in-group solidarity around goals that promise revenge and justice against perceived oppressors. McCauley and Moskalenko (2008) further describe the mechanism through which radicalization can result from discrimination along ethnic, racial, religious, national, or sectarian lines, and how this out-grouping draws individuals into extremist groups seeking to spark inter-group conflict – with extremists showering the in-group with praises of ideological purity and superiority over others.

Using this profile of radicalization, we may imagine the type of individual that is vulnerable to indoctrination. Individuals marginalized within their society, exposed to the trauma of becoming social pariahs or violence, who rally around causes associated with the social cleavage of differentiating features that separate them from those from whom they suffer, are likely to be vulnerable to indoctrinating strategies that leverage an aligning ideological point of view. Thus, in the context of increasing secularization, government oppression, wartime violence, and pluralistic integration of societies, we might imagine that individuals pushing back against these forces by diving deeper into a radicalized set of beliefs in their religious in-group may find themselves drawn to media that engages them on this dimension.

3.5.2 Co-optation

Instead of shifting a target audience's beliefs, a strategy of co-optation seeks to shift a target audience's expectations of future social, political, or economic benefits without shifting their ideological perspective. This strategy focuses on the substantive gains that an audience can obtain by supporting the extremist cause without emphasizing the ideology backing it. By promising a unique set of benefits afforded by extremists, this strategy leverages the present or future division of spoils obtained by extremists to secure recruits or funds. Through the lens of media strategy, co-optation corresponds to the use of material goods, services, and administrative and governance benefits provided by an extremist group in exchange for support. Thus, content reflecting co-optation should carry forward the thread of a strategy of engagement that uses depictions of these governance and market benefits to convince potential followers of the rewards of supporting extremists' causes irrespective of – or even despite – differences in ideological beliefs.

Co-optation belies a contractarian understanding of what extremists are seeking to do when they seek to take over territory and "act like a state": in exchange for access to a share of the fruits of citizens' labor in the form of taxation, governing authorities spend a portion of these rents to provide citizens of the state with public services – from security, in the most basic sense, to market subsidies, protections, and social welfare benefits in the broadest (Hobbes, 1671; Olson, 1993). It also belies an understanding of the extremist rebels-as-state as resulting from market failures that prevent individuals from engaging in beneficial economic exchange (Smith, 1776; Wintrobe, 2012), which, in turn, prevents the state from collecting an optimal share of that wealth. Extremists may not be operating with these particular scholars in mind (or perhaps they are – see Derluguian, 2004), but they are operating with similar rent-maximizing ideals in mind: in order to extract resources from people over time, resources that will almost certainly help extremists win on the

battlefield, extremist rebels understand that they must provide a sense of reasonable exchange, security, and potential violence for opting out of government rule and coming to the extremist's side of the conflict (Huang, 2016). A media strategy of co-optation therefore provides an opportunity for rebels to communicate this commitment ahead of time – particularly in the face of whatever prior reputation such a group carries from past experiences.

3.5.3 Intimidation

A messaging strategy of intimidation relies on the threat of punishment, rather than the promise of ideological, social, or material fulfillment, to convince a target audience to submit to extremist demands or get out of their way. This strategy uses the threat of physical, social, or psychological violence to deter its target audience from taking actions against extremists and encourage them to comply with extremist demands. Successful strategies of intimidation coerce audiences not only to refrain from acting against extremists, but to actively comply with their demands, whether that be staying home during an election, paying tribute or taxes to their leaders, dissolving local resistance, or avoiding extremists on the battlefield. Successful strategies of intimidation may also attract common enemies of a particular opponent to the extremist cause.

The critical component of a messaging strategy of intimidation is the depiction of violent coercion – the ability for an extremist group to employ violence, or the threat of violence, to force their audience to do what they want them to do. The use of violent coercion as a means for non-state political actors to intimidate their audiences into support, compliance, or deference is a well-documented phenomenon. It is perhaps best documented on the battlefield, where insurgent and combatant violence against the military forces of the state and opposing groups serves the primary function of resistance and competition for political control at the heart of the rebel cause. Off of

the battlefield, where violence is the primary tool of success, rebel actors may use violence against non-combatants strategically to improve member cohesion (Cohen, 2011; 2013), to signal resolve in pursuit of a military or political cause (Fearon, 1996; Kydd & Walter 2006), to signal ideological commitment to a cause or set of causes (Walter 2017), or to outbid competing organizations seeking to overshadow their influence in a crowded ideological space (Kydd & Walter, 2006). Rebel groups may also lose control of violence committed by their members in ways that compromise their strategic goals for any number of reasons, from spiraling crises of group formation and organization (Weinstein, 2006) to group fractionalization and within-group conflict (Lidow, 2017).

Messaging strategies of violence seek to realign an audience's understanding of an extremist group's use of violence to the strategic operational goals of that group at a particular time and, if possible, at a particular place. Extremists may not be able to control all expressions of violence committed by members of their groups. They may, indeed, also use violence for a multitude of purposes beyond that which they communicate via their messaging strategy. However, extremists can turn to messaging strategies of intimidation to shape the way their audience interprets the violence they see or hear about and transform the way that these events influence their actions. In doing so, messaging strategies of intimidation allow extremists to consolidate and concentrate the message they want to send coercively and downplay the potentially unintended consequences of violence committed either outside of the group's direct influence or outside the ideologically acceptable or consistent guiding principles of the organization's professed values or beliefs. In the language of violence, messaging strategies allow extremists to be very specific about the audiences against whom they threaten violence, and the scope of the

violence aimed at that audience, and the degree of violence facing that audience should they not comply with extremist demands or expectations.

3.6 Connecting Strategies to Audiences

Identifying a target audience is critical in determining what messaging strategy extremists will pursue. Audiences may vary on many underlying dimensions, but one compelling and intuitive dimension of variation is sympathy to the ideological cause of extremists. Sympathetic audiences are particularly prone to ideologically charged messaging because they are predisposed to adopt the beliefs of extremists based on their own prior ideological orientation. Malleable audiences are not particularly ideologically motivated, but may instead be co-opted by material, social, or political inducements communicated through a strategy of co-optation. Oppositional audiences are not susceptible to indoctrination or co-optation, but may bow to the threat of violence under a strategy of intimidation and give in to extremist demands. Identifying which audience extremists are engaging allows them to make a stronger, better tailored case for why individuals in that audience should join their group, give them money, or pass along information on their enemies.

We may use these underlying features of an audience to theorize how extremists might seek to target their media. Sympathetic audiences may be ideal targets for indoctrinating media, where nurturing further support for the ideological cause may bring individuals closer to providing support regardless of (or in addition to) the material or governance (co-optation) benefits advertised to them. Malleable audiences, on the other hand, may not be ideally responsive to indoctrination strategies because they lack the predisposition to accept extremist beliefs, but might instead be convinced to support the group that provides them the most compelling material, social, and governance support. Outside of these groups, oppositional audiences may be unpersuaded by

either indoctrination or co-optation strategies, but may be ideal targets for intimidation messaging that forces them to think twice before resisting extremists or even passively standing in their way.

However, identifying these audiences can be challenging for at least two reasons. First, appearing sympathetic to extremist goals or open to supporting extremists is usually costly: individuals who stick out as potential extremists or extremist supporters usually face reprisal from states, communities, and even family members who do not share their views, and this may push them to misrepresent their ideological beliefs and make them difficult for extremists to directly target. Second, extremists typically lack sufficient technology, local information, or political reach to individually sort people into these categories. While extremists (and mainline politicians alike) would prefer to rely on individual indicators of preferences when crafting their messaging strategies, they seem to lack granular information on the individuals that might consume their information. They also typically lack the funds and personnel to engage in personal outreach at such a large scale. These obstacles make it challenging for extremists to pinpoint the underlying sympathy of most individuals that they would like to reach.

To solve this problem, I argue that extremists rely on more easily identifiable identity cues to inform them of the likely orientation of their audience toward the group. This allows extremists to better tailor their messaging content to the *likely* sympathy of their audience based on an easily identifiable feature of their identity. Extremists can, in other words, develop a messaging campaign targeting English-speaking followers that differs substantially from that of the messaging campaign targeting Arabic-speaking followers. They can launch local messaging campaigns in areas they control that differ substantially from those they direct at foreign audiences thousands of miles away. Extremists may market themselves as predominantly ideological purists to one audience, promising radical policy support to one audience, while simultaneously marketing

themselves as benign providers of public services to another, asserting that they can outcompete the government in the provision of goods and services. Over time, extremists can shift these strategies across and within target audiences – relying on broad target audiences united by ethnicity, language, or sectarian or national identity as heuristics for groups of varying underlying sympathy to determine the game they play with that audience to secure their support or compliance.

Bringing interests, strategies, and target audiences provides us with a unified framework of extremist messaging operations. Extremists seeking to obtain recruits, resources, information, acquiescence, or deference from a particular audience first draw inferences from the broad, observable characteristics of that audience – relying on characteristics like geography, language, and online presence -- to determine the likely underlying sympathy of individuals within that audience. They then select messaging strategies – indoctrination, co-optation, or intimidation – that they believe most likely to get what they need from that audience. These strategies are dynamic; they change over time and across audiences depending on shifts in their priorities on the ground. As extremist's operational needs change, so does their use of messaging to persuade people to help them accomplish their goals.

3.7 The Operational Determinants of Messaging Strategies

Expanding on the framework introduced above, I argue that extremists tailor their messaging strategies to shifting operational priorities on the ground. Extremists conducting small-scale attacks at a local level have extraordinarily different operational needs than organizations actively conquering and maintaining control over territory or organizations seeking to build a competitive political coalition. These operational needs can push extremists to shift the audience

from which they seek recruits, funds, and compliance. These shifts in target audience correspond with changes in the strategy that extremists employ to gain access to the things they need.

I identify two shifts in operational priorities that explain the substantial variation in extremist messaging content across audiences. The first is a shift from insurgency and covert operations toward territorial control. In the course of a group's lifetime, extremists can transition from covert operations to overt contestation of territorial control against a government. The transformation of extremists from organizations built around secrecy and hiding to organizations that conquer and maintain control of territory out in the open requires extremists to seek support and compliance from audiences outside of those sympathetic to their cause in order to draw the recruits and resources necessary to effectively govern. The second is a shift from fighting the government to seeking to participate in it. Changes in the political system, changes in leadership, or changes in public support can change the political landscape within which extremists operate, leading them to enter the political system.

The shifts in audiences in both cases arises from a shift in recruitment, fundraising, informational, and compliance needs by rebels when transitioning from the smaller-scale operations of covert insurgency to overt territorial control or political participation. Covert insurgency requires a small group of trusted individuals to maintain secrecy while aiding small-scale operations (Shapiro, 2015; Berman et al., 2019). Co-optation can threaten the integrity of small scale operations that demand few personnel and secretive planning and whose management may be hindered or threatened by individuals that aren't ideologically invested.

Territorial control and political participation both require larger groups of people and resources to cooperate with extremists out in the open. This poses a challenge for extremists relying

on a predominantly sympathetic base of true believers: in order to successfully control territory or participate in a larger political system, they need to expand their base of support beyond their existing ranks. However, attracting people outside of the ideological rank-and-file requires a balancing act for these groups that forces them to reach beyond the radically charged messaging that secured their base. Under the framework introduced in this paper, I argue that extremists should employ a strategy of co-optation to draw less ideologically committed people to help them maintain control over territory or win political office. As extremists seek control over territory or political office that allows them access to greater political, social, and economic resources, extremists gain the ability to leverage present and future material awards that allow them to co-opt people without moving them ideologically.

This doesn't mean that groups should abandon their efforts to indoctrinate others entirely when they extend control over audiences outside of their rank-and-file. The inner workings of extremist organizations should still rely on the same ideological rigidity and discipline that defined the group prior to the shift, and extremists will still seek to indoctrinate others to their cause to maintain their ranks of true believers. However, the relative effort that extremists put into indoctrination should decrease as the needs of the organization expand beyond their ability to satisfy them with true believers alone – i.e., as extremists seek more buy-in from people outside of sympathetic audiences.

While these shifts in extremists' ground game both encourage a relative shift in messaging toward co-optation, the shift in strategies of intimidation in both cases is more nuanced. Territorial control exposes extremists to a new type of opposition: non-military civilians, sometimes cooperating with extremists' state or non-state opponents, that may resist extremist control by passing information to their enemies (Kalyvas, 2008), openly revolting against them (Arjona,

2016; Balcells, 2010) or shirking payments of taxes or tribute (Revkin, 2019). Political participation mobilizes civilian opponents in the existing political sphere to mount campaigns to prevent extremist entry or take down extremist momentum within the political system. Extremists can respond with a messaging strategy designed to intimidate these opponents into submission without risking widespread violence that could damage their standing as a governing force or political competitor. Extremists may still be forced to carry out violence to make these threats more credible in some cases, but the dissemination of messaging increases the scope of compliance for those intimidated enough to back down without calling extremists out.

In particular, when employed simultaneously with a strategy of co-optation, intimidation messaging differs from that employed by extremists operating outside of ambitions of territorial control or political participation. Under these circumstances, territorial control or political participation discipline extremists' threats of violence – limiting extremists' advertising of violence toward civilians, in particular, in areas under their control or influence to targeted, specific cases of opposition, rebellion, or subversion. The reasoning for this is simple: wanton violence against civilians while extremists seek to extract resources, recruits, or political influence from a wider audience of political supporters might alienate malleable individuals who would otherwise be co-opted by the group. A strategy of intimidation that does not adapt to shifts toward territorial control or political participation risks undercutting a successful strategy of co-optation. This doesn't, of course, mean that extremists don't carry out brutal attacks against civilians while pursuing these strategies. It just means that they do not promote these attacks as transparently in their messaging – preferring instead messaging that focuses more clearly on communicating the enforcement of their established order.

3.8 Targeting Messaging Strategies during a Shift

How do extremists target audiences with different messaging strategies? Extremists face a tactical dilemma in distributing their messaging: while individually targeted or precise group-targeted messaging strategies would be the most effective in targeting distinct audiences with tailored messaging conforming to distinct strategies, extremists typically lack the means – financial and informational – to target their audiences directly, particularly in the early stages of group formation (Weinstein, 2006; Staniland, 2012). Directly targeting an audience also carries substantial risks: keeping records of known or suspected supporters, informants, or compliers may imperil this audience if these records are seized, and a direct informational connection to a large audience may very well risk leaving the extremist group vulnerable to counterinsurgency measures that track their operations or distribution network (Shapiro, 2011).

Faced with limited resources and considerable risks to precise targeting of audiences, extremists turn instead to indirect targeting using proxies for the likely underlying sympathy of their audience. I identify and expand upon two dimensions here that extremists may use as proxies for targeting their messaging strategy based on the underlying sympathy of their audience: geography and language. I then illustrate how extremists may direct their messaging strategies through these targeting lenses in the empirical chapters to follow.

3.8.1 Geographic Targeting

Geographical targeting allows extremists to split messaging into geographic units – differentiating the message they send to and from a given area from the messages sent to and from another. By targeting individuals based on their areas in which they reside, extremists rely on both the geographic nature of political control and the geographic sorting of individuals by religion,

religious sect, culture, and nationality to target their messaging to individuals within these groups that are likely to be alike on an underlying dimension of sympathy for the extremist cause.

Using geography as a proxy for the underlying sympathy of an extremist audience carries two distinct advantages. First, it allows extremists to match their messaging strategies to their operational strategy – directly aligning their messaging with their military or administrative moves within discrete units of geography. Second, geographic targeting allows extremists to take advantage of the significant geographic sorting that occurs in multi-ethnic, multi-religious, and multi-sectarian societies – relying on a more easily differentiable proxy for the ethnic, religious, or sectarian identity of individuals than may otherwise be available by directly confronting individuals.

However, using geography as a proxy for underlying sympathy carries at least two major disadvantages. First, focusing on audiences based on where they are located misses substantial variation of individual sympathy to the extremist's cause within that geographical area. Individuals located in an area inhabited predominantly by co-ethnics or co-sectarians that have experienced hardship at the hands of the government may serve as the foundation for the geographic targeting of a strategy of co-optation in this area, but that message may fail to land among individuals in that region that strongly oppose the group in principle, or who are ideologically drawn to the organization. Second, extremists may be forced to use geographic proxies that do not line up well to divisions in the likely audience of their messaging based on pre-existing operational or political constraints. Thus, they may target messaging too broadly by adopting strategies that alienate wide swaths of their audience in a particular area – or wide swaths of the audience that is invested in that region – by folding them into the same strategic considerations of another audience within that geographic space.

3.8.2 Linguistic Targeting

Linguistic targeting allows extremists to split messaging into language-specific units, instead – differentiating the messages they send to audiences speaking a particular language from the messages they send to audiences speaking another. By splitting messaging among individuals in the same linguistic group, extremists can pursue different media strategies to different groups that may not be expected to understand one another or tune into media produced in another language. The underlying assumption made when targeting messaging by language is that a group can communicate distinct messages by producing content in different languages.

Extremists may operate under this assumption and target their messaging strategies to different linguistic groups in a couple of different ways. They may seek to use linguistic divisions to target distinct audiences more or less likely to be sympathetic to the extremist cause. One may think of this as a "one message, one group" strategy. For example, extremists may communicate a hardline ideological message to religious separatists in their Russian language propaganda at the same time that most of their Arabic language content features more governance-oriented releases, as was the case of Islamic State propaganda targeting separatist cells in the caucuses while seeking to manage newly acquired territory in Syria and Iraq in the early months of 2015 (Zelin, 2016). On the other hand, extremists may seek to pursue the same messaging strategy to different groups of individuals while harboring resentment for the other group within their language-specific releases. We might call this a "one message, many groups" strategy. Extremists may seek to communicate the same message to many groups across linguistic lines (lines that often overlap with cultural, ethnic, and religious divisions) while vilifying ethno-linguistic groups to one another, as again the Islamic State frequently does among non-Arabic speaking radical followers

overseas in its European-language releases and Arabic-speaking radical followers in its Arabic releases.

Using linguistic targeting carries significant limitations and risks. First among these is overlap in understanding and uptake of messaging among individuals in different linguistic target groups. Individuals may observe differences in the messaging across media content released in different languages and discount the authenticity, veracity, or integrity of the organization. Second, linguistic grouping of individuals as an approximation of underlying sympathy – or the population most likely to help extremists accomplish their goals in a particular context – may turn out to be a poor proxy. Targeting Russian separatists with ideologically charged indoctrinating messaging to attract the most radicalized of separatist leaders may alienate a critical audience of less hardline Russian speakers who would have been better served by a strategy of co-optation. Finally, extremists may fail to communicate consistently and effectively in the language used in their media, as al-Qaeda is widely acknowledged – and often ridiculed – for doing in their attempt at German language messaging in the first decade of the twentieth century (Mehrzad, 2010).

3.8.3 The Internet and Changing Messaging Capabilities

While the strategic framework introduced above alongside these targeting scopes may fit equally with in-person and online strategies of outreach toward potential followers (and deterrence of potential opponents), we have important reasons to believe that access to online media changes the way that extremists communicate with their followers. I present two here that are particularly relevant to this work.

First, the internet increasingly shrinks the world – increasing individual's awareness of phenomena far away from their place of origin (Xenos & Moy, 2004; Greschke, 2012) – but access

to the internet *also* increases individual's knowledge of and engagement in local social and political affairs (Norris, 2001; Smith et al., 2009), and not just in high-tech societies, but also those with relatively new access to internet technologies (Kent Jennings & Zeitner, 2003; Yue et al., 2019; Cho & Pelter, 2020). Thus, while the internet can greatly expand the audiences that extremists can reach with their media apparatus, it may also serve to amplify their engagement on the local level – using increasingly local footage to engage not just with people far away, but also with people closer to the areas in which they are active to communicate more broadly at a far lower cost than engagement with physical media or in person.

Second, the internet increasingly empowers extremist groups to reach out to diasporic communities without investing heavily in identification and in-person infiltration of their communities abroad. Contrasting with the recruitment dilemmas of the past wherein extremists would spend months-to-years infiltrating communities of migrants to engage them with the issues of their homelands hundreds-to-thousands of miles away (Heger et al .,2012; Gates & Nordas, 2015), extremists now have access to virtual communities of individuals vulnerable to marginalization in their first or second generation communities away from their ancestral homes and potentially more vulnerable to the influence of the messaging strategies described above. Several scholars have already demonstrated the influence of the internet in motivating individuals to pledge support for causes or travel to conflict zones (Mitts, 2019; Mitts et al., 2022).

Thus, we know that the internet can allow extremists to reach interested parties both locally and those connected to conflicts and causes in far-away places in a far less costly, far more accessible way. The internet may create a broader reach, but at the end of the day, its influence may be harnessed to fill the needs of organizations locally with similar ease.

3.9 The Shadow of Messaging Strategies

So far, I have introduced a framework that takes as given the strategic goals and needs of an extremist organization, evaluates the potential audiences from which extremists could extract their needs to accomplish their goals, and provides three distinct messaging strategies that they could direct toward these audiences to convince or compel them to do so. I have also provided an overview of the more practical targeting strategies that extremists might employ to target different audiences in the absence of clear cleavages upon which to divide their messaging. This allows us to theoretically trace the origins, dimensions, variation, target, and targeting strategy of extremist messaging. But what about its longevity?

In this section, I propose that precedents (and successive changes) in extremist messaging strategies cast a long shadow – locking extremists into a particular strategy among targeted audiences even after the strategic incentives that caused that shift have faded. I argue that this occurs for (at least) two reasons. First, extremists rely on the consistency of their messaging campaigns to communicate commitment to their goals with a target audience. Deviating sharply and significantly from a particular strategy may damage their relationship with potentially supporters, or embolden detractors inside and outside that audience to label them hypocrites. Second, extremists experience an operational delay in shifting messaging resources from one strategy to another. Shifting the physical, creative, and human capital inputs of a messaging strategy from one to another takes careful planning – a process that groups like the Islamic State actually seem to plan out in considerable detail internally. I unpack the foundations of these reasons before expanding on their consequences below.

3.9.1 Consistency as Commitment: Building a Stable Reputation

In order to build a stable foundation of support among audiences in volatile political environments, political opportunists like the extremist groups addressed in this dissertation cling to messaging consistency to develop lasting reputations among audiences unsure of their political surroundings. Political scientists have long observed reputation-building among rebel groups and fledgling governments alike in unstable political environments. Political actors may favor stable, consistent policies over more short-term-optimal choices in unstable, post-conflict environments to develop a reputation for resolve and a commitment to political consistency (Walter, 2009). Emerging state leaders may, in turn, develop incentives to stand by suboptimal policy choices due to the reputational costs of defecting built in by prior decisions (North & Weingast, 1989; Root, 1989, North, 1993; Fearon, 1998, Rosendorff, 2015).

Unlike state leaders, who have the machinery of the state at their helm to lock in their commitment to policy choices, extremist groups rely on messaging to reinforce consistency in their preferred governance and policy choices. Lacking a track record of consistent action across areas they contest, but do not yet control, their microphone becomes their commitment mechanism – with consistency in messaging at least partially replacing consistency in policy record in the construction of a reputation among prospective supporters. This conceptualization of commitment fits the socio-political characterization of commitment as "social capital" used to develop a trustworthy reputation with an uncertain audience (Wolfinger, 1960; Purdue, 2001; Gibler, 2008). Without a record to which to be held accountable, extremists hone in on consistency of messaging to win over audiences uncertain of their commitment to their claimed causes, tactics, and goals.

3.9.2 Logistical Obstacles to Sharp Changes

While commitment as a reputation-building mechanism forms the foundation of the stickiness of extremist messaging, it isn't the only obstacle that extremists face in sharply shifting their media strategy. An additional obstacle lies in the actual machinery of media efforts, particularly for media involving coordinated local video footage. Milton (2015; 2018) reveals the complexity of media strategy planning within the Islamic State – pointing to documents seized in raids of Islamic State storehouses showing plans covering months of groundwork, footage, execution, editing, and distribution going into the video series distributed online and on the ground by the organization over a two year period. Groups seeking to compete with the higher quality efforts of organizations like the Islamic State in the region have similarly stepped up their media strategies, and similar seizures of internal organization have revealed similar organizational obstacles to sudden shifts in media strategies in al-Qaeda in the Arabian Peninsula and the Taliban.

Investing in high-quality messaging strategies may force extremist organizations to lock themselves into production infrastructure and schedules unable to respond to quick changes in priorities, and this may cast a longer shadow on their messaging even when priorities in operations on the ground shift beneath them. Together with commitment-based reputation building, these obstacles create both incentives to maintaining a slower-moving messaging machine and obstacles to rapid shifts even when doing so may benefit the organization.

3.10 Main Hypotheses

Based on the above, I propose the following set of main hypotheses. In the empirical tests that follow, I test these hypotheses systematically:

A shift in extremist group operational strategy from small-scale operations to territorial contestation and control will cause:

1. A shift in messaging strategy
    a. toward strategies of *co-optation* of potential supporters outside the group's ideological base, and
    b. away from indoctrination of sympathizers;
2. A large and significant increase in the use of strategies of intimidation specifically targeting civilian resistors to extremist governance and control. A shift away from territorial contestation and control back to small-scale operations, on the other hand, will cause:
3. A *delayed* shift in messaging strategy away from a strategy of co-optation and toward a strategy of indoctrination due to the local reputational costs and logistical costs of sudden shifts in media strategy described above.
4. A *delayed* shift in messaging strategy away from a strategy of intimidation targeting civilian resistors to extremist control and toward a strategy targeting enemy combatants.

3.11 Data

To test these hypotheses, I exploit an explosion of data on the Islamic State following a near-decade-long series of military and cyber-operations to counter radical Islamist influence worldwide. I combine an unprecedented amount of online media data, geo-located event data on conflict operations conducted by government and intervening coalition forces across the Middle East, and historical sources on the demographic features, past political affiliations, and histories of violent oppression across several countries in the Middle East and North Africa within my analysis.

I consolidate these into a working data set that I use to test the hypotheses presented in the last chapter in the empirical chapters to follow.

3.11.1 Case Selection: Why the Islamic State

The Islamic State presents four opportunities that few other groups can offer to scholars studying extremist propaganda, governance, and territorial control. Together, these four opportunities make the Islamic State an ideal case for this analysis. The first is that the Islamic State is one of the few groups to successfully capture territory for a prolonged period of time in the past two decades – and it did so not just in one country, but in four (Liveumap, 2020; UN Habitat , 2016, 2018). One prominent feature of civil conflict in the 21st century is that the preponderance of these conflicts are asymmetric in nature (Berman & Matanock, 2015; Berman et al., 2018), meaning that the state dominates rebels in military capability. This usually leads to less effective rebel control over territory. In about 80% of active civil conflicts in the past twenty years, rebels have failed to secure control over territory for longer than one month (see ACLED, 2019; Huang, 2017). The Islamic State is one of a few prominent exceptions to this rule, but also stands out because of the variation in when and where it contested and controlled territory. The Islamic State began as a subordinated group within the al-Qaeda franchise without immediate territorial ambitions (Mendelsohn, 2018), and has clearly defined periods where it neither controlled significant amounts of territory nor wanted to. In addition, the Islamic State and affiliated organizations controlled territory on two different continents across four different countries at the height of its power, reaching into Libya and Nigeria in Africa while maintaining territory in Iraq, Syria, Egypt, and Afghanistan between 2014 and 2017. This variation of control over time and over geographic space provides ample variation to test the effect of territorial ambitions,

contestation, control, and loss on rebel media strategy while holding several group-specific variables constant.

Second, the Islamic State relied not just on informal authority in areas under its control, but on a well-publicized and decentralized structure of public administration that sought to provide services, enforce social policies, and collect taxes from people. Like other rebel groups, the Islamic State faced a choice when wresting control of territory: they could impose their own authority to govern, or defer to civilians to administer themselves (Arjona, 2016). In some, but not all, cases of IS control, the Islamic State sought to assert its own authority as an administrative power – directly participating in the formulation of policy and the collection and distribution of resources in places it got a foothold (Revkin, 2020). Variation on this dimension across areas under Islamic State control allows me to test hypotheses about the incentive to govern on the production of rebel media while holding several organizational characteristics constant.

Third, the Islamic State developed an expansive media presence that paralleled its expansion of territorial conquest and control. This media presence, while loosely centralized under a media production wing, allowed for the independent production and distribution at the local level under territory claimed, contested, and controlled by the Islamic State (Milton, 2016). The sheer availability of content allows me to test theories about when and where rebel groups use media as a local support strategy without relying heavily on quantitative measures that buckle under the weight of low sample sizes.

Finally, extensive policy and media coverage of the Islamic State's rise and fall by not just local, but also international, sources facilitated the collection of much more data than would typically be collected on the rebel groups' structure, operations, and media production. This means

that we know not only that the Islamic State produced (and continues to produce) a lot of media, but we can be reasonably confident that this media is catalogued and survives as it makes its way through several channels of communication – not only from the Islamic State to its target audience, but also through the data collection processes of private companies (The Direct Method, n.d.), scholars of rebel media (Zelin, 2015; Mitts, 2019; Mitts et al,. 2020), and opposition military researchers (Milton, 2016, 2018). Primary source clearinghouses like jihadology.net, Twitter activists like the Anonymous community, and internet archiving services like Archive.org allow users to identify, isolate, and save content before making it accessible to those studying extremists, rebels, and propaganda. This allows us some assurances about the survivability of rebel media data that we do not have in smaller conflicts that run below the public policy radar of western states where information may not be tracked, stored, and accessed later by scholars.

3.11.2 Media in the Islamic State

For the dependent variable, rebel media strategy, I draw on media data predominantly collected from scholar of jihadist movements Aaron Zelin's primary source clearinghouse jihadology.net, which seeks to collect primary source propaganda materials from Salafi-Jihadist groups online and is used in several cutting-edge empirical studies of jihadist propaganda (Zelin, 2015; Walter & Phillips, 2019; Mitts et al., 2020). I scrape this source for all media produced between 2011, the earliest date of release available for the Islamic State, the primary subject of my within-case analysis, and 2018, the point at which most key organizations substantially slow the distribution of propaganda through online channels due to increased online censorship on social media – pushing much of their distribution of media content from public channels to private channels that limit the empirical scope of inference (Twitter, 2018; Mitts, 2019).

I focus the data collection efforts and quantitative analyses on media produced by the Islamic State between 2014 and 2018. I rely on *this* data from *this* range of time for three reasons. First, the richness of this data – over 5,600 independent video, audio, and periodical releases distributed and available through publicly accessible channels over five years – allows me to conduct a fine-grained analysis of media behavior not possible with other groups. Second, the decentralized structure of the Islamic State's media bureaucracy allows me to identify the region of origin of media releases with a higher degree of certainty than possible with other groups in my data, which in turn allows me to explore hypotheses concerning the geography of media production and distribution. Third, this range of time allows me to focus my analysis on the *public* dissemination of rebel media because prior to late 2017, groups like the Islamic State were still able to disseminate propaganda content easily on social media hubs like Facebook and Twitter without specific pushback from the companies themselves (Mitts, 2019).

The unit of analysis in this paper – the parameter that describes what each observation in my dataset used for analysis represents – is the governate-month. That means that every observation represents the media produced by the Islamic state in a particular governate (officially designated region) under their control in a roughly 28-31 day period characterized by months in the Islamic calendar. I have chosen this unit of analysis for a simple reason: it best reflects what we know of the way that the Islamic State organizes its media production and release. Scholars focusing on how groups like the Islamic State produce propaganda have provided us with internal documents recovered in areas formerly under Islamic State control that show how they planned the production and dissemination of media, and these have revealed that the Islamic State plans their production and distribution schedule on weekly and monthly calendars (Milton, 2018a,

2018b; Zelin, 2015). I can thus reasonably capture the data generating process – the way the data is *actually* generated by the source – by using this unit of analysis.

3.12 Measuring Thematic Content in Media

Detecting thematic content in a corpus of thousands of text, audio, and video releases poses a challenge to researchers. Relying on human coders to read, listen to, or watch each release requires potentially hundreds of labor hours, and in the context of rebel media, may expose human coders to potentially traumatic content. In order to identify thematic appeals within rebel media without relying exclusively on human coders, I turn to computational tools – IBM's Watson transliteration service and Amazon's Rekognition object identification API – that allow me to strip each type of release into a bag of words and images that allows coders to classify the thematic content of each release without directly observing the media (Mitts, Phillips, & Walter 2020). Watson transliterations turn every word spoken in a video into text. Rekognition identifies each object within a $1/10^{th}$ second frame of a video.

Raw Release

IBM Watson Transliteration

وهي بازدياد وانظمة اه صحيح ابو رائد مهما كان بس ليلة عنيز بيل سبيكس رجب
بلديات استريبو مشترك تميزت بريسينو لما غيرنا استاذ بريش ومن روبروا يوصل
كيكا فريخ حتى استريز امامنا كافكا زيباكويرا اشتركت فيها لأن بودوي والكرواتي
افترار يستثني بيروت تجري وشن يجد كان ردود وغيرها ونستورد بيصير غير...

Amazon Rekognition Output

Aircraft	Bomb	Chain Saw	Fire Screen	Kicking
Ammunition	Bomber	Cuff	Fist	Knife
Archery	Bonfire	Cutlery	Flame	Lava
Armor	Bow	Dagger	Flare	Machine Gun

Figure 3: Three sources of information available to coders when deciding which content category to code a particular media release.

I employ supervised machine learning to code media themes across my media corpus. This process comprises three steps. First, human coders observe the title, Watson transliteration, and the most prevalent unique objects in a sample of 500 releases and observe the accompanying title banner (example above under "Raw Release") before determining the significant themes present based on a list of themes (adapted from Mitts et al., 2020; full list of themes and coding protocol provided in data appendix). This adapted list of themes separates content into four categories based on each feature's theoretical closeness to the theme described in the theory section above: indoctrination, co-optation, intimidation, and an "other" category representing content that does not fit into these thematic categories. Specific coding criteria, examples of releases in each category, and the descriptive statistics for this process can be found in the Coding section in the accompanying Data Appendix.

Next, I use two separate machine learning algorithms – a naïve Bayes classifier and a support vector machine – to classify the data not coded by human coders. All analysis below is robust to both classifiers, and the accuracy, precision, and recall of each classifier are provided in the accompanying data appendix under Table A. The specific coding instructions for classifying media into thematic content categories can be found in the Coding section of the Data Appendix below.

As a secondary measurement, I also employ a dictionary method of classification to isolate governance media within my corpus. To do this, I use the same human coding from the supervised learning protocol to isolate a dictionary of "thematic terms" – terms associated with indoctrination, co-optation, and intimidation terms associated with order-preserving violence, violence against civilians, and "oppositional violence" terms associated with content depicting extreme violence employed against an enemy. I then classify the remaining releases as containing each type of media if their titles contain these terms. This coding produces strikingly similar results to the supervised learning results detailed above. I provide these results of these comparisons in Figure 1 in the Data Appendix.

3.13 Measuring Territorial Control

Rebel territorial control – and extremist control in particular – is difficult to measure at a granular level. Scholars and policymakers alike typically rely on intelligence and journalistic accounts, which come in the form of event data that must be aggregated to provide useful geographic information. To measure this across space and over time, I use geolocated event data tracking Islamic State activity provided by geolocated event data from LiveUMap, a service endorsed by several organizations, including UN Habitat and the World Bank, to construct a

governate-week and governate-month measure that indicates whether or not the group was reported to control or contest any territory in the region. LiveUMap aggregates reports from credible, verifiable local and international sources on territorial changes in control and tracks these changes in maps it publishes and releases to scholars and policymakers. Since LiveUMap provides information on a daily basis and maintains a geographic account of territory under group control, I aggregate this data to the week and Islamic month level to match it with my media data. I then combined this with data from an emerging dataset on the breadth and depth of Islamic State control across Iraq, Syria, and Libya to produce a governance-level measure of territorial control within that territory at any time at the week- and month- level.[2]

3.13.1 Lagging Control

Including geographic control of a territory as a variable introduces a critical question about the timing of control's effect on media strategies: how long after the Islamic State establishes geographic control over a territory should we observe an effect on media strategy? In the analysis below, I draw on insights from Milton et al. (2018a; 2018b), who find in documents seized from the Islamic State's internal media production unit that the process takes about one month to begin the type of high-quality local production characteristic of their online propaganda disseminations when they establish a stable presence in a location. I therefore conduct my analysis with a one-month lag to reflect this. However, in order to account for the potential that this lag may be an arbitrary feature of my design choice, I display the results of the panel matching analysis with both no lag (t) and lags of up to six months out (t + 6, where each unit represents a month).

---

[2] A special thank you to Thomas Nemecheck for excellent research assistance in collecting this data.

3.13.2 Measuring Control Variables

In addition to the key independent variables that follow from my hypotheses, I draw on several demographic and geographic features of the people and places included in this study as control variables in my regression models. In the primary statistical models evaluated in chapters 4 and 5, I include five control variables (all at the subnational governate level): proportion Shi'a Muslim, religious minority presence, history of violent government oppression, estimated household income, presence of natural resources, and governate population as a proportion of total country population.

Proportion Shi'a Muslim measures the proportion of each governate that identifies as Shi'a Muslim. Shi'a Islam is one of two dominant sects of Islam, and is the majority sect of three countries in the Middle East and Central Asia: Iraq, Iran, and Bahrain. Significant Shi'a populations also exist in Syria, Lebanon, Libya, and Somalia. The Islamic State adheres to an extreme, fundamentalist take on Islam that excludes Shi'a their conceptualization of Muslims, thus allowing them to justify their killing, maltreatment, and enslavement under the principle of *takfir* (Maher, 2016). The inclusion of *Religious Minority Presence* as a control variable captures a similar distinction, as the principle of *takfir* also typically applies to these populations. A significant population of Shi'a Muslims or a populous religious minority may affect the Islamic State's willingness to provide or advertise the provision of public services in a particular governate, and should therefore be included in this analysis. I draw data on the proportion of Shi'a Muslims in each governate and the presence of a significant religious minority from the most recent census data available in each governate.

History of Violent Government Oppression measures the number of incidences of state-sponsored violence that have occurred in a governate. State sponsored violence may traumatize civilians, leaving them more open to alternatives to a government they have been conditioned to fear and mistrust. I measure this variable by taking the count of all incidents of state sponsored violence from the UCDP georeferenced event dataset (GED) and aggregating them by governate (Sundberg & Melander, 2013). This is not a perfect measurement by any means of the imagination (see Shaver et al., 2020 working paper), but it does represent the best data we have to date on these events.

Presence of natural resources captures whether or not a region contains lootable natural resources that rebels can produce and sell to fund their operations. Natural resources can greatly affect how rebels and states alike govern territory (Ross, 2015) and how they treat civilians (Weinstein, 2006). We might also think that these effects might impact how rebels broadcast their willingness and capability to govern. I therefore include a variable indicating the presence of a lootable natural resource that rebels can access in my analysis.

3.14 Empirical Design & Results

3.14.1 Regression Analysis

In order to ascertain the influence of territorial control on rebel media strategy, I employ a regression analysis designed to isolate the correlation between territorial control and media strategy while controlling for the effects of several potentially confounding variables described in more detail above.

I employ three distinct types of regressions on three different dependent variables – proportion of indoctrination media, co-optation media, and civilian-oriented intimidation media out of total media released in each period – in my first round of analysis: an ordinary least squares regression, a one-way fixed-effects regression accounting for the potentially confounding effects of unaccounted factors that vary over time ("time-variant" features), and a two-way fixed-effects regression accounting for time-variant and geographic-variant features. Fixed effects regressions are particularly useful on cross-sectional and panel data in contexts in which the researcher does not have access to reasonable amount of data on all potentially relevant control variables, but does have enough across-unit (geographic, in this case) and over-time data to construct a model that reasonably estimates and accounts for confounders that vary across these units as *fixed* (Greene, 2011; 2017). I also include various controls in each of these model specifications, as reported above.

Because each of these regressions is based on an OLS estimator, and I measure media strategy as the proportion of each thematic category present in that unit's total media, we can interpret the results with relative substantive ease. Accounting for the varying controls and fixed effects included in each model, we can say that territorial control by an extremist group in a given governate in a given month corresponds to:

1. A roughly 73%, 18%, or 30% *increase* in the proportion of local media releases reflecting a theme of **co-optation** respectively. Each of these results is statistically significant at the $p < 0.01$ level;
2. A roughly 42%, 37%, or 22% *decrease* in the proportion of local media releases reflecting a theme of **indoctrination** respectively. While the OLS model with controls is statistically

significant at the $p < 0.01$ level, both fixed effects models are significant instead at the $p < 0.05$ level;

3. A roughly 17%, 11%, or 7% *increase* in the proportion of local media releases reflecting a theme of **civilian intimidation** respectively. While the OLS model with controls is statistically significant at the $p < 0.01$ level, the month fixed effects models is significant instead at the $p < 0.05$ level, while the two-way fixed effects model is significant only at the $p < 0.1$ level.

Each of these three model specifications across each dependent variable above comes with notable limitations. Most notably, none of these three models allows me to causally estimate the effect of territorial control on extremist media strategy. While my two-way fixed effects model backs a more compelling argument for a rigorous statistical relationship between these two variables, recent scholarship has thrown into question the ability of even this type of regression analysis to identify causal relationships in even the best panel data (Imai & Kim, 2020). In addition, each of these tests effectively tests the potentially path-dependent effect of territorial control on media strategy after control has been achieved or lost. This may call into question what the actual effect of taking control or losing control over territory is on media strategy in subsequent time periods, excluding periods where control does not change hands. It may also beg the role of reverse causality – namely, whether media strategy actually causes territorial control, and thus accounts for the effect estimated in this model.

3.14.2 Panel Matching Analysis

I take two steps to strengthen my analytical argument that these correlations provide meaningful and significant evidence of a causally relevant relationship. First, I have lagged my

independent variable of interest – effectively testing the effect of territorial control *one time period before* the media strategy of a given period. For example, each regression will be estimating the effect of control at time *t-1* (think: October) on media strategy in time *t* (think: November). While the effectiveness of lagged variables in reducing time-based confounding in regression analysis is not absolute by any means, there remains a robust literature supporting the usefulness of this tool in slightly strengthening our causal interpretation of regression output by eliminating the immediate possibility of temporal reverse causality, i.e. media strategy in a given time period causing territorial control, which would be a valid theoretical concern in this context (although see Bellemare et al., 2017 for an expanded discussion of why these methods are insufficient).

Second, to address the threat to inference posed by two-way fixed effect model specifications, I have employed panel matching using the PanelMatch package developed by Imai, Kim, & Wang (2020) to provide more robust evidence of a causal relationship between extremist control over territory and media strategy. By seeking to approximate an experimental design by matching features of units across time and the assignment of "treatment status" of the primary independent variable of concern, panel matching eliminates some of the confounding effects of repeated time-invariant observations that do not take into consideration the time varying treatment assignment of units over time (Imai et al., 2021). This process allows me to better identify the causal effect of extremists taking over territory in a particular governate on the local media strategy that they employ in subsequent months in that same governate by comparing that governate's media strategy to a basket of other governates with matched political, demographic and economic covariate features.

I present the results of this procedure across my three dependent variables of interest in Figures 4-6 below..

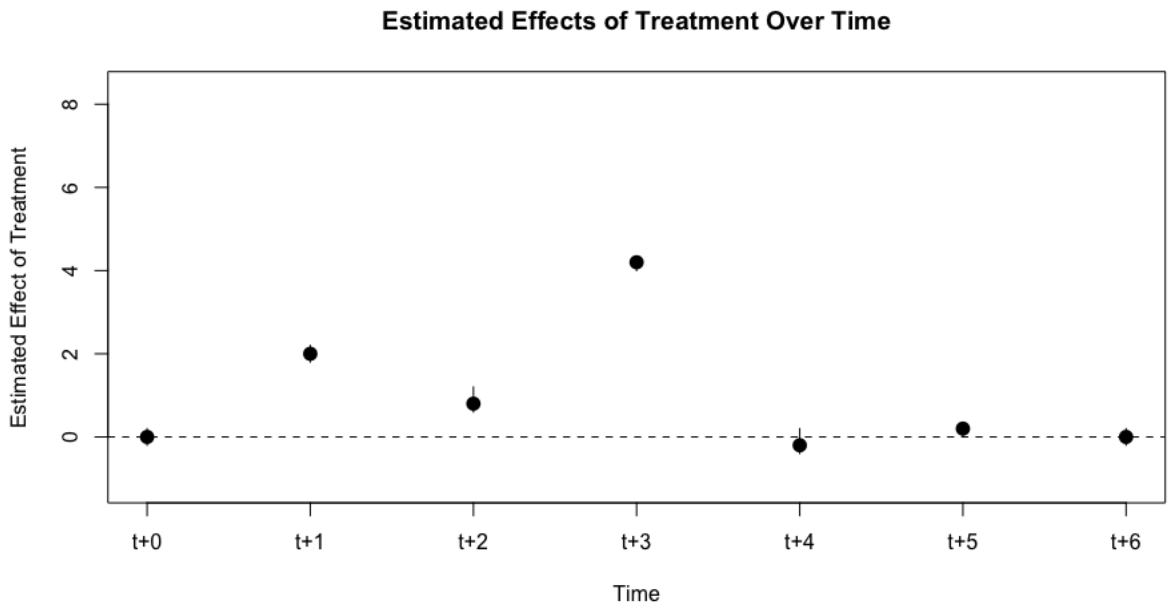


Figure 4: Panel Matched Analysis of Effect of Territorial Control on Proportion Co-optation

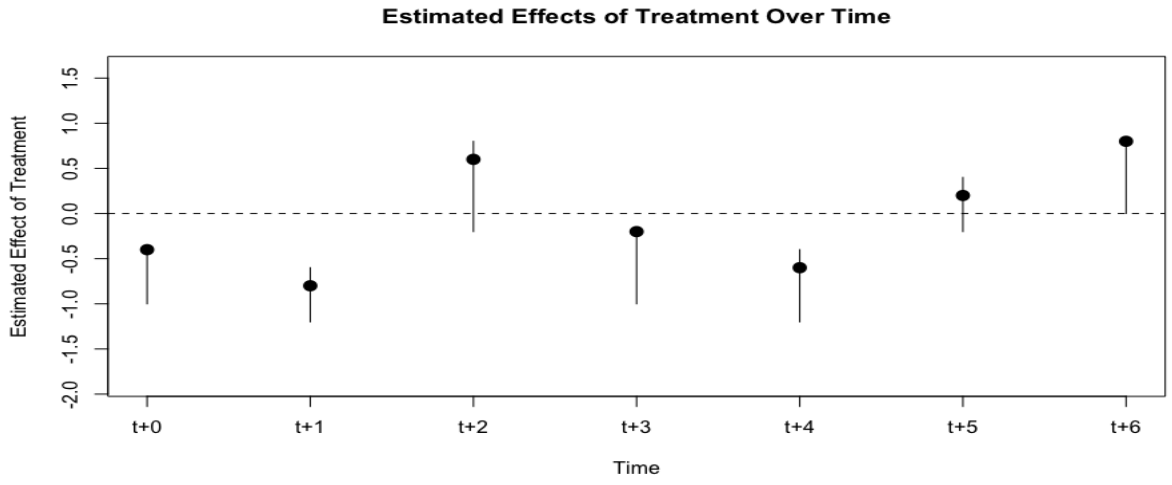


Figure 5:Panel Matched Analysis of Effect of Territorial Control on Proportion Indoctrination

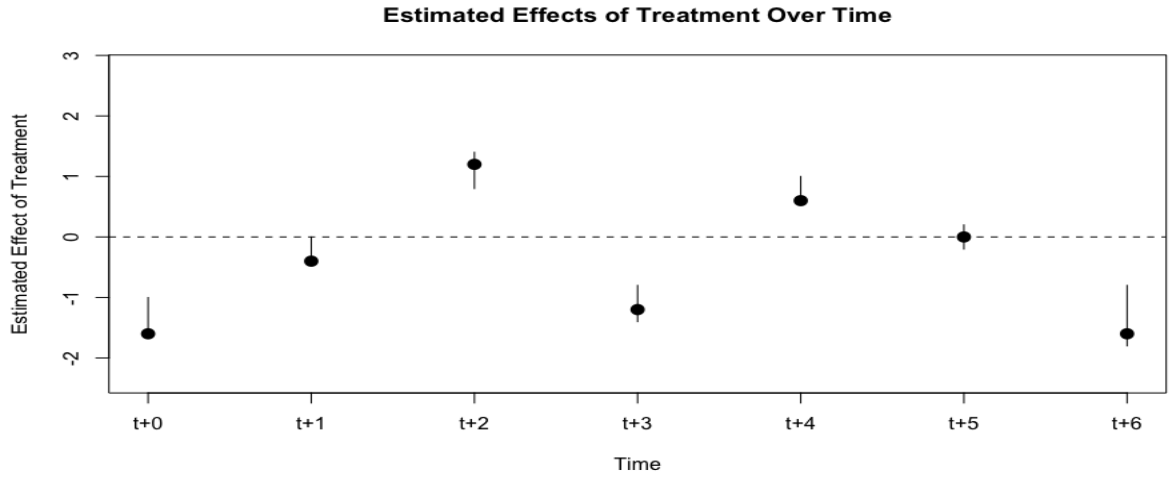


Figure 6: Panel Matched Analysis of Effect of Territorial Control on Proportion Civilian Intimidation

The panel matched results are consistent with the unmatched regression estimations above, with the onset of territorial control corresponding to:

1. Strong and significant ($p < 0.05$) increases in the proportion of co-optation-themed releases in Islamic State media for each of the three months following it;
2. Strong and significant ($p < 0.05$) decreases in the proportion of indoctrination-themed releases in Islamic State media for each of the three months following it;
3. A more tenuous, less robust, and less consistently significant relationship with civilian intimidation media across the time period following it.

3.14.3 Evaluating the Hypotheses

This empirical evidence allows me to evaluate the hypotheses driving this analysis. Taken together, I draw three conclusions. First, I find universally strong and statistically significant evidence that territorial control by the Islamic State shifted the group's messaging strategy toward a strategy of co-optation (Hypothesis 1a). Second, I find universally strong and statistically

significant evidence that territorial control by the Islamic State shifted the group's messaging strategy away from a strategy of indoctrination (Hypothesis 1b). Third, I find some preliminary signs but do not find consistent nor robust evidence that territorial control by the Islamic State shifted the group's messaging strategy toward a strategy of civilian intimidation (Hypothesis 2).

3.15 Conclusion & Discussion

How does extremist messaging adapt to changing priorities in the everchanging battle for civilian hearts and minds? In this paper, I have argued that extremist messaging is not rigid and strictly ideological, but rather locally adaptable to shifting operational priorities in the organization's broader group ambitions. I have brought an expansive collection of data to this question – applying cutting-edge tools in the processing and analysis of text- and video- as data. Finally, I have conducted an extensive array of rigorous empirical tests that provide robust and significant evidence for the relationship between territorial control and a shift in rebel media strategy away from indoctrinating content and toward broader messages of co-optation.

This paper makes two central contributions to the broader field of political science. First, on the subject of political communication in conflict, this paper introduces a novel conceptualization and operationalization of extremist messaging strategies – building on prior conceptualizations of propaganda, radicalization, and state-building to construct a framework of strategic interactions that lend themselves to implications tested above. Second, on the issue of extremism and radicalization, this paper builds upon other recent works (Walter & Phillips, 2020; Mitts et al., 2020) to expand the paradigm of extremist group recruitment and organization beyond the paradigm of radicalization, ideological appeals, and violence – showing instead that extremist groups respond to operational incentives to highlight less ideologically charged and violent

components of their organization and goals when their local priorities shift to control and governance. Together, these contributions illuminate a path forward as extremist communications continue to expand in reach, volume, and importance in an increasingly interconnected world.

3.16 Acknowledgements

Chapter 3, Mainstreaming Extremism, is also solo authored. This paper originated from a recurring question I had while working with Prof. Barbara Walter on our project investigating which extremist groups use online propaganda in Iraq and Syria: given that a group decides to adopt online propaganda as a tool, what strategic logic might extremists employ when determining what kind of content to disseminate to which groups of people? This paper represents my best answer to this question. I humbly acknowledge the employment and mentorship/collaborative support of Prof. Barbara Walter in collecting the data used in the empirical analysis of this paper, as well as the countless hours spent in her office digging deeper, and deeper, and yet deeper into how to properly articulate the research question and dig down into the theory of the paper. I also humbly acknowledge Prof. David Lake's contribution through a similarly long stretch of office hours probing the three core strategies that emerge at the heart of this paper. Finally, I acknowledge the intellectual contribution of Prof. Tamar Mitts, whose feedback throughout our time as collaborators was instrumental in shaping my thinking on this topic.

3.17 References

Albert, K. (2022). What is rebel governance? Introducing a new dataset on rebel institutions, 1945–2012. *Journal Of Peace Research*, 002234332110518. https://doi.org/10.1177/00223433211051848

Arjona, A. (2016). *Rebelocracy: Social Order in the Colombian Civil War* (1st ed.). Cambridge University Press.

Bellemare, M., Masaki, T., & Pepinsky, T. (2017). Lagged Explanatory Variables and the Estimation of Causal Effect. *The Journal Of Politics*, *79*(3), 949-963. https://doi.org/10.1086/690946

Bergen, P. (2001). *Holy War, Inc.: Inside the Secret World of Osama bin Laden* (1st ed.). Free Press.

Berman, E. (2011). *Radical, Religious, and Violent: The New Economics of Terrorism* (1st ed., pp. 1-314). The MIT Press.

Berman, E., & Matanock, A. (2015). The Empiricists' Insurgency. *Annual Review Of Political Science*, *18*(1), 443-464. https://doi.org/10.1146/annurev-polisci-082312-124553

Berman, E., Felter, J., Shapiro, J., & McIntyre, V. (2018). *Small Wars, Big Data: The Information Revolution in Modern Conflict* (1st ed., pp. 1-408). Princeton University Press.

Bott, C., Dickens, R., Moffitt, J., Smith, E., Rowley, T., Lark, R., & Thompson, G. (2009). The Internet as a Terrorist Tool for Recruitment & Radicalization of Youth. *Homeland Security Institute Publication Number RP08-03.02*, 17-01.

Doob, L. (1935). *Propaganda: Its Psychology and Technique* (1st ed., pp. 1-424). Oxford.

Doob, L. (1950). Goebbels' Principles of Propaganda. *Public Opinion Quarterly*, *14*(3), 419. https://doi.org/10.1086/266211

Fortna, V. (2015). Do Terrorists Win? Rebels' Use of Terrorism and Civil War Outcomes. *International Organization*, *69*(3), 519-556. https://doi.org/10.1017/s0020818315000089

Greene, W. (2017). *Econometric Analysis* (1st ed.). Pearson.

Horgan, J. (2008). From Profiles to Pathways and Roots to Routes: Perspectives from Psychology on Radicalization into Terrorism. *The ANNALS Of The American Academy Of Political And Social Science*, *618*(1), 80-94. https://doi.org/10.1177/0002716208317539

Huang, R. (2016). *The Wartime Origins of Democratization: Civil War, Rebel Governance, and Political Regimes.* (1st ed.). Cambridge University Press.

Imai, K., & Kim, I. S. (2020). On the use of two-way fixed effects regression models for causal inference with panel data. *Political Analysis*, *29*(3), 405-415.

Imai, K., Kim, I., & Wang, E. (2021). Matching Methods for Causal Inference with Time-Series Cross-Sectional Data. *American Journal Of Political Science*. https://doi.org/10.1111/ajps.12685

Kilcullen, D. (2015). *Out of the mountains: The coming age of the urban guerrilla*. Oxford University Press.

Maher, S. (2016). *Salafi-Jihadism: The History of an Idea*. 1st Edition. Oxford: Oxford University Press.

McCauley, C., & Moskalenko, S. (2008). Mechanisms of Political Radicalization: Pathways Toward Terrorism. *Terrorism And Political Violence*, *20*(3), 415-433. https://doi.org/10.1080/09546550802073367

Milton, D. (2018). *Down, but Not Out: An Updated Examination of the Islamic States Visual Propaganda*. Combating Terrorism Center West Point United States.

Milton, D. (2018). *Pulling Back the Curtain: An Inside Look at the Islamic State's Media Organization*. Combating Terrorism Center West Point United States.

Mitts, T., Gregoire, P., & Barbara, F. Extremist Radicalization and Recruitment in an Internet Age. *Revise And Resubmit At Journal Of Politics*, (Working Paper).

Nemechek, T., and Gregoire P. (n.d.). "Conceptualizing Rebel Territorial Control in the Age of Contestation."

Olson, M. (1993). Dictatorship, Democracy, and Development. *American Political Science Review*, *87*(3), 567-576. https://doi.org/10.2307/2938736

Revkin, M. (2019). Relative legitimacy and displacement decisions during Rebel Governance. 1-36. *http://dx.doi.org/10.2139/ssrn.3365503.*

Revkin, M. (2020). What Explains Taxation by Resource-Rich Rebels? Evidence from the Islamic State in Syria. *Journal of Politics*, *82*(2). https://doi.org/10.1086/706597

Roeder, P. G. (2018). *National secession: Persuasion and violence in independence campaigns*. Cornell University Press.

Sanín, F., & Wood, E. (2014). Ideology in civil war. *Journal Of Peace Research*, *51*(2), 213-226. https://doi.org/10.1177/0022343313514073

Sanín, F., & Wood, E. (2014). Ideology in civil war. *Journal Of Peace Research*, *51*(2), 213-226. https://doi.org/10.1177/0022343313514073

Shapiro, J. (2015). *The Terrorist's Dilemma: Managing Violent Covert Organizations.* (1st ed.). Princeton University Press.

Shaver, A. (n.d.). Biased Media Reporting and Our Understanding of Global Conflict.

Smith, A. (2009). *The Wealth of Nations*. Blacksburg: Thrifty Books.

Sundberg, R., & Melander, E. (2013). Introducing the UCDP Georeferenced Event Dataset. *Journal Of Peace Research*, *50*(4), 523-532. https://doi.org/10.1177/0022343313484347

The Redirect Method. (n.d.). Moonshotteam. https://moonshotteam.com/the-redirect-method/

Walter, B. (2017). The Extremist's Advantage in Civil Wars. *International Security*, *42*(2), 7-39. https://doi.org/10.1162/isec_a_00292

Walter, B. F., and Gregoire P. (2018). Rebel Online Propaganda in an Internet Age. *Under Review*.

Weinstein, J.M. (2006). *Inside Rebellion: The Politics of Insurgent Violence*. 1 edition. Cambridge ; New York: Cambridge University Press.

Wiktorowicz, Q. (2005). A Genealogy of Radical Islam. *Studies In Conflict &Amp; Terrorism*, *28*(2), 75-97. https://doi.org/10.1080/10576100590905057

Wintrobe, R. (2012). Autocracy and coups d'etat. *Public Choice*, *152*(1-2), 115-130. https://doi.org/10.1007/s11127-011-9862-3

Wood, E. J. (2016). *The Wartime Origins of Democratization: Civil War, Rebel Governance, and Political Regimes*. New York: Cambridge University Press.

Zelin, A. Y. (2015). Picture or it didn't happen: A snapshot of the Islamic State's official media output. *Perspectives on Terrorism*, *9*(4), 85-97.

4. Propaganda and Radicalization in an Internet Age

Tamar Mitts Gregoire Phillips Barbara F. Walter

4.1 Abstract

Over the past decade, a large number of extremist and hate groups have turned to internet platforms to inspire mass violence. Currently, there is little reliable evidence on how such campaigns radicalize targeted audiences. This study illustrates an inexpensive, systematic, and scalable way for estimating the effect of online propaganda on the behavior of individuals on social media platforms. Our method utilizes several machine learning algorithms to detect recruitment messages in online propaganda, identify their dissemination on social media, and quantify how exposed users react. We apply this method to content produced by the Islamic State that was widely shared on Twitter between 2015 and 2016. We find that propaganda messages conveying the material, spiritual, and social benefits of joining ISIS were effective at increasing online support for the group, while content displaying brutal violence was ineffective at generating support across a wide range of videos. The study illustrates how powerful machine learning techniques can be applied to detect granular messages in online content, a method that can be replicated to study other information campaigns.

4.2 Introduction

One of the most unsettling phenomena emerging around the world is the rise of lone-wolf terrorist attacks perpetrated by citizens in their home countries. Since 2000, the number of domestic terror attacks in the United States has increased nearly threefold (US Dept. Homeland Security, 2019). About two thirds of these attacks have been committed by individuals who pledge

allegiance to far right nationalist groups; the remainder claim to be motivated by support for radical Islamist groups. These acts of violence have occurred in almost every state in the U.S. and the vast majority of perpetrators have been American citizens or legal residents. This includes Dylann Roof, a white supremacist who murdered nine African Americans at a church in Charleston, South Carolina, Omar Mateen a self-proclaimed "Mujahideen" who opened fire in a gay nightclub in Orlando, Florida, and John T. Earnest, a white nationalist who attacked a synagogue in Poway, California. Earnest claimed to be inspired by Australian-born terrorist Brenton Harrison Tarrant who live-streamed via Facebook his attack on a mosque in Christchurch New Zealand, and by Robert Bowers who had attacked a Pittsburgh synagogue the year before. These attacks are particularly disturbing because they are so difficult to stop; since September 11, no foreign terrorist organization has successfully launched a deadly attack on US soil;[3] all post 9/11 attacks have come from individuals radicalized from within.

Most policymakers believe that the internet and social media have played a critical role in radicalizing these individuals. Many extremist groups today – including the Islamic State, al Qaeda, as well as numerous Neo-Nazi groups – maintain public social media profiles to communicate with their followers. Anyone with access to the internet can now quickly find and follow any extremist group online: radical left-wing environmental groups, radical right wing anti-government groups, anti-abortion groups, white supremacist and anti-Semitic groups, and transnational Salafi-jihadist groups. The great fear is that these individuals will self-radicalize in the comfort of their homes and become motivated to launch mass violence.

3 Peter Bergen, Terrorism in America after 9/11. New America Foundation. 2020.

But does internet propaganda really have this effect? Scholars aren't sure. On one end of the debate are those who argue that individuals who choose to follow extremist groups on the internet have already become radicalized prior to following the group. Their radicalization process started long before, often due to pre-existing political, economic, social, or psychological conditions. It's not the online messages that made them more radical, but past events such as personal experience with discrimination or an earlier trauma (Horgan, 2012). On the other end are those who argue that the internet is a highly effective way to radicalize individuals, as it provides the conversations and the community of like-minded individuals that move people further down the path to extreme beliefs (Horgan, 2008).

The evidence in favor of either argument remains anecdotal at best. Currently, there is little systematic evidence on whether online propaganda produced by terrorist groups has a causal role on inspiring support for extremism. We also have no systematic evidence for what types of messages, if any, are more or less successful in radicalizing targeted audiences. It is no surprise, therefore, that the effectiveness of counter-messaging campaigns has been, at best, hit-or-miss (Jackson and Costello, 2019).

In what follows, we introduce a scalable method for estimating the impact of online propaganda on the behavior of individuals on online platforms. Our method utilizes several machine learning algorithms to detect recruitment messages in audio and video files produced by violent groups, identify their dissemination on social media, and quantify the reactions of exposed users. By applying this method to Islamic State propaganda, we provide, to our knowledge, the first large-scale empirical evidence of the impact of ISIS propaganda on individuals who followed the group online.

Our method involves several steps. First, we applied deep learning algorithms to detect objects and transcribe audio in ISIS propaganda videos. We used the output of these models to identify themes or 'topics' in online content produced by the group. Second, we detected the dissemination of ISIS propaganda on Twitter by tracking mentions of each release in the text of Twitter posts. We focus on Twitter because it was widely used by the Islamic State between 2015 and 2016—the years on which we focus in this study. Third, we identified the set of users who were exposed to these materials, as well as users who were not, by drawing on publicly available network data from Twitter. In the fourth step, we used several machine learning models to detect expressions of support for ISIS in millions of Twitter posts. Finally, we estimated a wide range of high-frequency Difference-in-Differences models that compared content produced by exposed followers on Twitter to content produced by non-exposed followers in the day before and after each propaganda dissemination.

Our analysis reveals several important findings. First, propaganda messages distributed by ISIS inspired higher levels of support among the group's online followers. Individuals exposed to propaganda on Twitter were more likely to convey their sympathy with the ISIS, talk about foreign fighters, and discuss the group's actions in Syria and Iraq. Second, certain propaganda messages were much more influential than others. Propaganda content showing the material benefits of joining ISIS, kinship among the group's soldiers, as well as content speaking to grievances and ideology, strongly increased endorsement for the group.

Contrary to popular expectation, however, propaganda depicting brutal violence did not generate mass support for the Islamic State on Twitter. Users who followed accounts that disseminated videos showing atrocities decreased expressions of sympathy with the group in the aftermath of exposure. In fact, the same messages that increased support for ISIS – such as those

talking about kinship or rewards – became ineffective when they were coupled with violent imagery in the videos. Exposure to brutalities led the group's online followers to post less pro-ISIS tweets even months after violent propaganda appeared on their timelines.

But not all followers had this negative response. In fact, a small minority of Islamic State followers on Twitter had the opposite reaction. Hardcore ISIS supporters who expressed particularly high levels of sympathy with the group *before* exposure to propaganda increased their support for ISIS after seeing its brutalities. This was true for all kinds of violent content, both general militant scenes, as well as extreme violence showing executions. This pattern corroborates the intuition underlying many models of radicalization – that individuals who move further down the path of extremism become more tolerant to (or even supportive of) brutal violence.

To our knowledge, this is the first study to provide systematic, large scale, micro evidence on the effect of Islamic State propaganda on Internet followers around the world. Our findings suggest that what attracted most online support for ISIS was not the violent propaganda that made the group so famous, but messages that conveyed the material, social, and spiritual benefits of recruitment.

4.3 Five Mechanisms of Radicalization

Five main mechanisms have been offered in the social science literature for why an individual would choose to embrace a radical group. People may come to endorse such a group because they are (1) dissatisfied with their political, economic, or social status, (2) offered material or personal rewards for support, (3) converted to the cause via ideological indoctrination, (4) interested in joining the strongest or most committed group to a cause, or because they are (5) seeking a sense of community and kinship. Below, we summarize each of these causal logics and

use them as the basis of a series of empirical tests on the types of online content that extremist groups might produce when targeting potential supporters on social media.

4.3.1 Individual and Group-Based Grievances

Scholars have long recognized the role that discontent and suffering may play in leading individuals to support extreme ideas (Weinstein, 2008; McCauley & Moskalenki, 2008; Carter et al., 2014; Bass, 2014; Lyons-Padilla et al., 2015). Individuals may be radicalized by personal experiences with poverty, anger over unequal economic and social opportunities, personal abuse at the hands of state officials, or hurtful encounters with ethnic, racial, or religious discrimination (Slootman et al., 2016; Dalgaard-Nielsen, 2010). Immigrants to Europe from Muslim majority countries, for example, have long faced multiple barriers to assimilation, are poorer on average than the native population, and are increasingly the target of hostility from far-right political parties (Burrows, 2016; Jamal, 2008; Karam, 2012; Naber, 2008; Stack, 2015; Mitts, 2019). Each of these factors could cause an individual to reject the status quo and seek to replace it with a new and more radical type of system.

Individuals may also come to support an extremist group because of abuse and mistreatment directed at their identity group by those in power (Sageman, 2004; Marx & Scott, 1976; Wickham-Crowley, 1992; Gurr, 1971). For example, in 2017, a British-born citizen launched an attack against pedestrians near Parliament because he was outraged by the British government's decision to fight Muslims in Syria and Iraq. The Islamic State has capitalized on this sentiment by creating hundreds of hours of videos showing how the problems of potential recruits are connected to the larger causes for which ISIS is fighting (Penman, 2015). If individuals are radicalized, in part, because they believe that they or their group are victims of abuse or

discrimination, then online messaging that emphasizes real or imagined wrongs should increase their support.

4.3.2 Ideological Indoctrination

An equally plausible path to radicalization is the power of ideas. Numerous studies have revealed that individuals can be converted to a cause through careful ideological indoctrination (LoCiero & Sinclair, 2007; Van Ijzendoorn, 1997; Flaherty, 2003; Zimbardo, 2007). ISIS recruiters, for example, consistently pushed the message that going to Syria was a spiritual obligation of Muslims around the world (Callimachi, 2017). Radical groups can disseminate information online that slowly inculcates audiences into the beliefs and aspirations of their movement, and systematically and uncritically accept extreme worldviews (Dandachli, 2017; Gómez et al., 2017; Wood, 2015). They can also share members' personal joy at conversion and their delight at their new life. If ideological indoctrination is at the heart of online radicalization, then internet messages designed to educate or reinforce a particular set of beliefs should be especially successful.

4.3.3 Material and Personal Inducements

The third mechanism is more instrumental. Individuals may come to support extremist groups because they believe they will gain material or personal rewards as a result. Historically, violent groups have tried to recruit individuals by offering compensation including money, social services, club goods and insurance, the promise of marriage partners, or offers of spiritual fulfillment or forgiveness (Popkin, 1979; Lichback, 1995; Berman, 2009; Berman & Laitin, 2008; Wood, 2003, Sageman, 2004; McCauley & Moskalenko, 2008; Horgan, 2008; Weimann, 2016; Bouchard, 2015; Rink & Sharma, 2018; Vidino & Hughes, 2015). For example, a young Scottish

woman who became an online recruiter for the Islamic State frequently wrote that joining ISIS included "a house with free electricity and water provided to you due to the Khilafah [the caliphate] and no rent included" (Fantz & Shubert, 2015). If individuals respond to material and personal rewards, then these types of internet messages could induce them to embrace an extreme group even if they, themselves, are more moderate.

4.3.3 Displays of Strength and Outbidding

An emerging literature in social psychology points to a fourth mechanism by which individuals can be radicalized. Human beings appear to have an innate desire to dominate other groups and may be hard-wired to identify with an organization perceived to be particularly powerful (Petersen, 2002; Tajfel & Turner, 1979; Pratto et al., 1994; Sidaneous & Pratto, 2001). The Islamic State, for example, often depicted itself as possessing "continuous momentum" and overpowering, crushing, or defeating its rivals (Callimachi, 2017). Individuals also seek to identify the group most dedicated to a cause and thus more likely to win a war (Kydd & Walter, 2006).[4] If people are motivated by a desire to be on the "winning team," then messages that showcase the military strength of a group and its willingness to use violence could be particularly effective.

4.3.4 Social Benefits and a Sense of Belonging

Finally, scholars have also found that individuals are drawn to organizations, in part, for a sense of kinship and belonging. Individuals may come to support a radical organization because they feel alienated socially and hope to reconstruct a lost identity; membership offers them a way

4 Bloom 2006; Kydd & Walter 2006

to connect to a network of like-minded people such as Gilles Kepel, Farhad Khosrokhavar, and Oliver Roy (Dalgaard-Nielsen 2010; Stanliland, 2011, 2014; Horowitz, 1985; Tajfel & Turner, 1979, 1986). Al-Qaeda supporters living in Western Europe, for example, appear to have been motivated to embrace extremist interpretations of Islam as a way "to recast and rationalize their sense of exclusion" (Dalgaard-Nielsen, 2010; Roy, 2004: 324). If this is true, then messages that emphasize community, comradery, or foster a sense of belonging should be particularly useful in gaining support (Hall 2015; Shane et al., 2015).

Each of these mechanisms identify different messages that may appeal to individuals drawn to extremist groups. Note that these mechanisms are not mutually exclusive. Group seeking to recruit via online means might include multiple different messages in a single video hoping that one or more will work. In the context of the Islamic State's online campaigns, a video may contain battle scenes in which ISIS is victorious together with stories about camaraderie and offers of marriage partners. There may also be good reasons to flood the internet with different messages knowing that different themes will appeal to different types of potential recruits. Ideological messages, for example, may draw in more spiritual individuals, while material inducements might appeal to more practical ones. It is also possible that an individual is influenced by each of these messages but at different stages in their life. For example, some people may first decide to follow ISIS because they experience a personal trauma – perhaps the unjust death of a close friend or relative, abuse at the hands of the state, or a brutal prison term. These experiences then make them more receptive to messages that emphasize group grievances, which then leads them to ideologically rich content which converts them to the cause. All of these mechanisms, therefore, may be instrumental in moving a person down the path of radicalization.

4.3.5 Scope Conditions

This brings us to the scope of this research. To make our analysis tractable and as close as possible to identifying the impact of propaganda, we focus on a very particular phase of a person's radicalization process.[5] In the context of online radicalization, this phase occurs after an individual has chosen to follow an extremist group on the internet, but before he or she engages in violence (if at all). The goal of this paper is to determine whether a person who actively follows the Islamic State online can be made even more extreme simply by receiving online propaganda and, if so, what messages are most likely to pull him or her closer to the group.

As a result, our study does not reveal what caused a person to become interested in the Islamic State in the first place. It could be that ISIS Twitter followers have been genetically pre-disposed to radical ideas, or have disproportionately suffered abuse, injustice, and discrimination at the hands of their government or society more generally (Horgan et al., 2016). It could also be that the individuals in our dataset have had a higher rate of mental illness, antisocial and paranoid behavior and lower self-esteem than the population as a whole (Nussio, 2017). Our study also does not address what ultimately convinces individuals to engage in violent acts, although some

---

5 There is no agreed definition of radicalization. We define radicalization as a "change in beliefs, feelings, and behaviors in directions that increasingly justify intergroup violence and demand sacrifice in defense of the ingroup." Definition comes from: Clark McCauley and Sophia Moskalenko, "Mechanisms of Political Radicalization: Pathways Toward Terrorism," Terrorism and Political Violence 20:415-433. 2008. P. 416. Jeff Victoroff "The Mind of the Terrorist: a Review and critique of psychological approaches," JCR, 49(1) 2005. 3-42. Alex P. Schmid, "Radicalisation, De-Radicalisation, counter-Radicalisation: A Conceptual Discussion and Literature Review," ICCT Research Paper, March 2013

suspect that the Islamic State has been attempting to groom online followers to become offline soldiers, once they became highly sympathetic to the cause. As prior research shows, some of these individuals eventually engaged in violence, but many did not (Neumann, 2013; de la Porta, 2018). What brings individuals online and what convinces them to engage in violence are likely to be driven by additional factors that demand their own distinct studies. The goal of this paper is to determine what types of online propaganda move individuals down the path of radicalization once they have decided to go online.

4.4 Data and Methods

In this study, we focus on propaganda disseminated by the Islamic State, for several reasons. First, ISIS has been one of the most active producers of online extremist content to date, providing us with ample data to study. Second, the group's media outreach served as a model for how to build a sophisticated online presence that has been copied by other groups around the world. If we want to design effective counter-extremism measures, we need to understand what 'worked' for one of the most prominent groups in this space.

We analyze ISIS propaganda disseminated between January 1, 2015 and December 31, 2016. We scraped these materials from *jihadology.net*, a widely-used archive which captured most, if not all, of the material officially disseminated by ISIS during this time. We limit our analysis to content that was released between 2015 and 2016 because comprehensive Twitter data – which we use to study reactions to ISIS propaganda – are not available for earlier years due to Twitter's

suspension of these accounts.6 In total, 1,557 disseminations by the Islamic State were archived during these years.[7]

In addition, we limit our analysis to audio and video materials. The vast majority (90%) of ISIS propaganda archived during this period was in video and audio format.8 Since our goal is to analyze the impact of the group's propaganda on the online behavior its Twitter followers, we prefer to focus on materials that were more likely to be consumed in real time.[9] Another benefit of focusing on audiovisual material is that users who are not literate in Arabic can still understand at least parts of it, making propaganda content more accessible to followers around the world. Below, we explain how we coded recruitment themes in ISIS propaganda audios and videos.

4.4.1 Coding ISIS Propaganda

Key to our analysis was our ability to determine what messages the Islamic State was sending in each propaganda file it disseminated. If individuals sitting at home clicked on a file they received from an ISIS account on Twitter, what types of material would they hear and see?

---

6 Islamic State accounts have been continually suspended from Twitter during the study's time period. To obtain information on accounts before they were suspended, we draw on data collected in real time. Since many accounts were suspended before the start of this data collection, information on the online activity of ISIS sympathizers prior to 2015 is sparser.

7 In the Jihadology archive, these materials were classified with "the-islamic-state" tag.

8 See Table 1 in the Supplementary Materials.

9 Audiovisual content tends to be consumed on social media at higher rates than articles. See, for example Templeman, Mike. "17 Stats And Facts Every Marketer Should Know About Video Marketing." Forbes, September 6, 2017. <https://www.forbes.com/sites/miketempleman/2017/09/06/17-stats-about-video-marketing>

4.4.1.1 Coding the Audio Content of ISIS Propaganda

Both audio and video propaganda files included voice messages. These came in different forms, such as people having conversations with each other, reading a text to the camera, singing, or giving speeches. We analyzed the content of these voice messages using a three-step process. First, we transcribed the audio into text using an automated transcription algorithm.[10] Then, we annotated a random sample of 400 transcripts to detect recruitment themes that related to the mechanisms discussed above.[11] Table 1 shows these themes and the broader recruitment categories to which they belong. For example, content depicting Western atrocities against Muslims or anti-Muslim hostility was coded as part of the 'grievances' category; content discussing public goods provided in ISIS's territories was coded as part of the 'material benefits' category; and content conveying religious observation or calls for jihad were coded as part of the 'ideology' category. The Supplementary Materials provide more information on the definitions of these categories, as well as several examples of coded propaganda transcripts.

Our third step was to scale up the propaganda coding using machine learning. We used the labeled propaganda files that we hand coded to train algorithms to detect the recruitment themes

10 Propaganda files were in mp3 and mp4 format. We used the Watson Speech to Text API in Arabic, which translates human voice into text by drawing on information on Arabic language structure and grammar as well as the composition of audio signals in each file. See IBM Cloud. "Speech to Text: About." November 12, 2018. URL: https://cloud.ibm.com/docs/services/speech-to-text?topic=speech-to-text-about#about.

11 The initial hand coding was done by Arabic-speaking individuals we solicited on Figure Eight, a crowdsourcing platform commonly used by academic researchers. Each theme in each propaganda transcription was coded by three coders, and labels were chosen based on majority agreement.

in the rest of the propaganda that was disseminated between 2015 and 2016. Specifically, we trained elastic-net generalized linear models – one for each recruitment theme – to predict the content of the remaining propaganda materials. We provide more details on the algorithms and their performance in the Supplementary Materials. To ensure accuracy, we also had research assistants qualitatively validate the machine coding and fix any errors.

We would like to emphasize that each propaganda file often included multiple messages that fit into different categories. That is, a single file could be coded as containing several different recruitment themes. One propaganda file, for example, started with a speech laying out prayers, then moved to encourage followers to carry out jihad while mentioning atrocities against Sunni Muslims, and finishing with a call to join the Islamic State. This file was coded for having the following themes: "Religious education and observation," "Call to jihad," and "Western atrocities against Muslims." [12] In our analysis, described Section V, we examine how recruitment themes that span over thousands of propaganda disseminations were received by the group's followers on Twitter.

4.4.1.2 Coding the Visual Content of ISIS Propaganda

In addition to audio content analysis, we also draw on recent advances in computer vision to uncover imagery in ISIS propaganda videos. The visual content of ISIS videos was particularly important because this is where ISIS often displayed military strength and brutality – a key measure of the *Dominance & Outbidding* mechanism. To code militant scenes, we use the Amazon Rekognition API, a visual analysis tool that uses convolutional neural networks to detect, frame-

12 See Supplementary Materials for details on this propaganda file and its coding.

by-frame, activities and objects in videos.[13] Amazon's algorithm draws on a large training set that includes both violent and non-violent imagery. In many commercial settings this tool is used for detecting 'unsafe' content; here, we use it to code militancy in ISIS propaganda videos without watching the videos ourselves. The algorithm detected 12,325,635 objects in 1,072 videos, with a total of 1,965 unique objects.[14]

Table 5: Themes in Islamic State Propaganda. The table displays different types of recruitment themes in Islamic State propaganda, along with the broader recruitment categories to which they belong. The Supplementary Materials provide more information on each category.

Broad category	Recruitment theme
Grievances	Western atrocities against Muslims
	Life in the West hostile to Muslims
	Sectarian appeals
	Celebrations of martyrdom
Ideological indoctrination	Religious education and observation
	Religious punishment
	Call to jihad
	Motivational speeches
Material benefits	Marriage and family life
	Public or club goods
	Life of an ISIS follower
Social benefits	Comradery between ISIS soldiers

13 See https://aws.amazon.com/rekognition/ for more information.

14 The algorithm provides a confidence level for detected objects, which ranges between 0 and 100. In our analysis, we examine reactions to videos with these objects; our results hold for various confidence levels (> 50, 60, 70, 80).

To identify videos depicting militancy and violence, we coded each object identified by the algorithm for whether or not it showed violent scenes, and found 76 such objects. To differentiate between different kinds of violence, we further divided this set of objects into four groups, as shown in Table 6. The first type of imagery includes air and artillery fire – bombs, airstrikes, and rockets, for example. The second includes scenes with general ruins (e.g., rubble, smoke, or fire) that can take place after explosions. The third category includes physical violence against humans, which we believe captures propaganda showing atrocities committed the group. Objects in this category include weapons, knives, swords, kicking, chains, and injury, among others. Finally, we examine a more general militancy category, consisting of scenes with military or soldiers. In total, these objects appeared 888,332 times in the videos.

In order to understand how these types of content were received by potential supporters, and whether violent propaganda was attractive to those who followed the group online, we needed information on the exact times in which propaganda messages were disseminated, as well as information on individuals who were exposed to them, and their various reactions. In the next section, we describe our rich Twitter data, which allows us to answer these questions.

Table 6: Objects Detected in Propaganda Videos that Depict Militancy and Violence. The table shows objects detected in Islamic State propaganda videos that appear in scenes showing militancy and violence.

(A) Air and artillery							
Cannon	Tank	Aircraft	Missile	Rocket	Warplane	Bomber	Bomb
Destroyer	Navy	Battleship	Mortar				
(B) Ruins and fire							
Smog	Smoke	Ruins	Demolition	Rubble	Fire	Funeral	Flare
Flame	Eruption	Tomb	Lava	Tombstone	Burner	Fire screen	Extinguishing fire
Skeleton	Forest fire	Bonfire	Camp fire				
(C) Violence against persons							
Weaponry	Weapon	Gun	Handgun	Cutlery	Ammunition	Armor	Arrowhead
Rifle	Whip	Sniper	Dungeon	Porthole	Bangles	Machine gun	Armory
Kicking	Prison	Shield	Boxer	Mallet	Bullfighting	Sword	Cuff
Blade	Dagger	Knife	Torch	Chain saw	Wedge	Archery	Bow
Fist	Chain	Injury	Hammer				
(D) Military and soldiers							
Army	Military	Officer	Soldier	Troop	Swat team	Police	Military uniform

4.4.1.3 Propaganda Dissemination on Twitter

Our next challenge was to identify who was exposed to these materials on Twitter. To identify propaganda disseminations, we draw on a large dataset of ISIS networks who were active on Twitter between 2015 and 2016 (Mitts, 2019). Using information on network structures (i.e., who followed whom), as well data on the historical timelines of these accounts (i.e., what users have posted over time), we examine how exposed followers reacted to various types of recruitment themes in ISIS online propaganda.

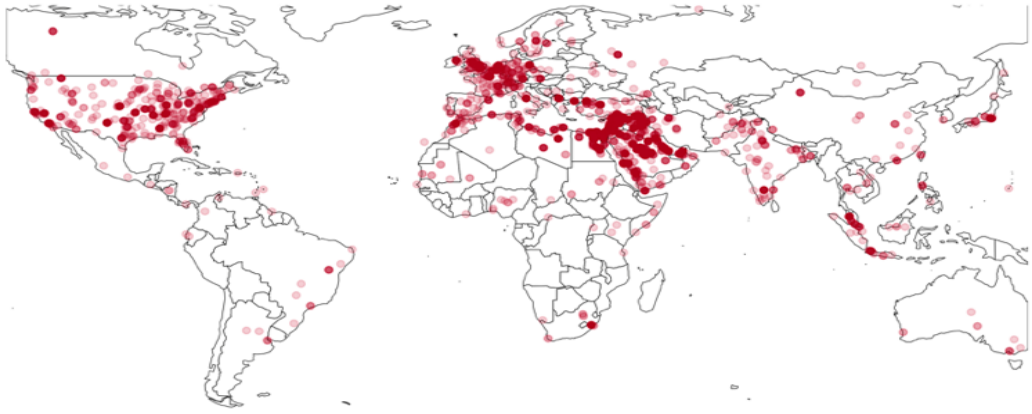
ISIS Twitter Data. Our Twitter data includes information on accounts that were identified by several anti-ISIS hacking groups for disseminating the group's propaganda on the platform, as well as their followers. Information on these accounts was collected in real time, before many of them were suspended from Twitter, and includes data on user profiles, locations, historical tweet

timelines, and lists of friends and followers. While accounts that disseminated propaganda had a clear connection to the organization, their followers were a more heterogeneous group, consisting of individuals who actively supported the organization, as well as individuals who were 'on the fence' or those that followed ISIS accounts for non-ideological reasons. Figure 1(a) shows the location of users who disseminated ISIS propaganda during the years of the study, and Figure 1(b) shows the location of their followers. The Supplementary Materials provide more information on these accounts, as well as more details the Twitter data collection.

Table 7: Locations of ISIS Propaganda Disseminators and Their Followers. Panel (a) shows the location of accounts that disseminated ISIS propaganda between 2015 and 2016. Panel (b) shows the location of their followers.

(a) Location of Propaganda Disseminators

(b) Location of Exposed Followers



Detecting Propaganda Disseminations. To detect propaganda disseminations, we searched for tweets that contained the titles of ISIS propaganda releases. Disseminating tweets

tended to mention the propaganda release and share a link to the video or audio file. Table 3 presents examples of anonymized propaganda disseminations shared on Twitter.[15]

15 Since the platform was actively suspending ISIS-affiliated accounts during this period, we were only able to capture a subset (34%) of the propaganda disseminated on Twitter during the years under review. To test whether the propaganda we found on Twitter is significantly different from the propaganda we were not able to capture, we examined variation between the content included in our dataset and the content that is not. Table B in the Supplementary Materials shows very little evidence of selection, where almost all content categories have identical proportion in both samples.

Table 8: Anonymized Twitter Disseminations of a Propaganda Video. The table shows three dissemination of one propaganda file that link to YouTube.

(a) Banner

Image Source: Jihadology.net

(b) Crowdsourced Transcriptions

الأسئلة بعد الجمع والتفريغ مع المسؤول العام لأجناد مصر# الحوار المفتوح
الحوار المفتوح مع المسؤول العام لأجناد مصر الأسئلة بعد الجمع والتفريغ وانتظروا اللقاء قريبا بإذن الله
إنتظروا اللقاء قريبا بإذن الله# الأسئلة بعد الجمع و التفريغ#مع المسؤول العام لأجناد مصر# الحوار المفتوح

(c) Common Words

مع اللقاء قريبا بإذن الله الحوار المفتوح المسؤول العام لأجناد مصر بعد الجمع

(d) Disseminations in ISIS Twitter Database

#أجناد_مصر# مع المسؤول العام لـ .. الحوار المفتوح#مؤسسة_الكتائب# http://t.co/yxHc1loYsH #مصر http://t.co/I2PBTfVcTG

أجناد مصر - الحوار المفتوح مع المسؤول العام: http://t.co/r6yejGRMop

أجناد مصر - الحوار المفتوح مع المسؤول http://t.co/k0EQW5X04q على @YouTube أعجبني فيديو

Identifying Twitter Users Who Were Exposed to ISIS Propaganda. Our next step was to identify who received the propaganda. When ISIS released a particular video file, which users were likely to see it? This is challenging to measure, as Islamic State sympathizers could have been exposed to the video in multiple channels. In this study, we draw on the *high frequency* nature of our data to measure instantaneous reactions of users to followed Twitter accounts that disseminated propaganda. Using information on network connections, we identified the followers of disseminating accounts and got their tweets in the few days before and after the dissemination.

To ensure that we are not picking up on spurious time trends, we compared these reactions to content posted by ISIS sympathizers on Twitter who did not follow the disseminators. To get the comparison group, we identified, for each dissemination, users that did not follow the disseminators but were similar to those who did. We identified the comparison group with nearest neighbor matching,[16] where we matched users on the basis of their levels of expressed sympathy with ISIS in the week prior to the propaganda release. After identifying a comparison group for each propaganda dissemination, we obtained these users' tweets in the days before and after it was shared on Twitter.

Measuring the reactions of ISIS Twitter followers. Our main dependent variable measures whether followers of ISIS expressed more sympathy with the group after they were exposed to a particular propaganda message. For this purpose, we created text-based measures capturing several pro-ISIS content categories: expressions of sympathy with the group, discourse on life in ISIS-controlled territories, mentions of foreign fighters, and communications about the group's actions in the Syrian civil war. We use a pro-ISIS index variable that sums the first three categories to capture more general pro-ISIS sentiment. In addition, we measured anti-West rhetoric, as such content frequently appeared in ISIS Twitter discourses (Cunningham et al., 2017; Mitts, 2019)

Since the volume of tweets was high, we used supervised machine learning to code the textual content of the tweets. We used crowdsourcing to hand-code a set of 30,000 randomly

16 Daniel E. Ho, Kosuke Imai, Gary King, Elizabeth A. Stuart (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistica*1*l*4*Software*, Vol. 42, No. 8, pp. 1-28. UR: http://www.jstatsoft.org/v42/i08/

selected tweets,[17] and used it to train models that predicted pro-ISIS rhetoric across more than five million unlabeled posts.[18] At the end of this process, each tweet received scores that reflected the extent to which it reflected pro-ISIS rhetoric.[19] These measures allow us to observe what ISIS followers on Twitter were posting on the platform each day, as well as how their content changed – if at all – after they were exposed to propaganda materials. We provide summary statistics and details on model performance in the Supplementary Materials.

4.5.2 Research Design

We use a Difference-in-Differences framework to determine whether followers of ISIS who were exposed to its propaganda were more likely to express sympathy for the group after

17 The tweets were in English, Arabic, French, and German. We used the crowdsourcing platform Figure Eight to annotate Twitter posts. See: https://www.figure-eight.com. Each tweet was labeled by three coders, and label(s) were chosen based on majority agreement – that is, at least two of the three coders had to assign the label to the tweet. We took several precautionary steps to reduce the likelihood that the human coders (971 in all) might inadvertently bias the coding of radical content. First, to be sure that the coding instructions were easy to understand, we confirmed that a student research assistant was also able to correctly code tweets using these instructions. Second, we randomly assigned each tweet to multiple coders, which should cause idiosyncratic biases from individual coders to cancel out on average. Third, we manually checked a random sample of coded tweets to ensure that the coding reflected the correct topics.

18 We used logit models with elastic-net regularization for each topic and language. The Supplementary Materials provide more details on the models and their performance.

19 We created binary versions of these scores, where tweets with scores above the mean were coded as mentioning a given theme, and tweets that below the mean were coded as not mentioning the theme. In addition, we created a pro-ISIS index variable that combined tweets expressing sympathy with ISIS, discussing on life in ISIS territories, and mentioning of foreign fighters.

exposure. We are able to do this because not all followers were exposed to every piece of content, allowing us to compare the tweets of exposed individuals to the tweets of users who were not. Our key identifying assumption is that in the absence of a propaganda dissemination, exposed and non-exposed followers would follow parallel trends in their online support for the Islamic State. That is, over-time changes in support for ISIS should not be significantly different between the groups before propaganda disseminations.

Figure 2 empirically tests this assumption. The Figure plots pre- and post- time trends in pro-ISIS tweets for the group of users who followed propaganda disseminators (red) and the group of users who did not (blue). The x-axis is the number of hours between the time of a propaganda dissemination and the time in which these users posted on Twitter. In order to observe time trends for all disseminations simultaneously, we normalized the difference in hours between the events and the timing of Twitter posts. The figure shows that the overtime trends in pro-ISIS rhetoric were parallel in the pre- treatment period. Only after propaganda disseminations we observe a shift in those trends, where pro-ISIS content by individuals who were exposed to propaganda increased, but the rhetoric of those who were not remained the same. The effect is strongest within the first 5-10 hours of propaganda disseminations.

To examine these patterns more systematically, we estimate the following tweet-level Difference-in-Differences model, where $Y_{i\,jk}$ measures pro-ISIS content in tweet *i* posted by user *j* before and after propaganda dissemination *k*:

$$Y_{ijk} = \beta_1 Exposed_{jk} + \beta_2 Post_{ijk} + \beta_3 (Exposed_{jk} \times Post_{ijk}) + \delta_k + \epsilon_{ijk} \qquad (1)$$

The independent variables are $Exposed_{jk}$, coded 1 for tweets by users who followed the disseminators of propaganda *k*, and 0 for used who did not; and $Post_{ijk}$, which is coded 0 for tweets

posted in the day before the dissemination and 1 within one day after. In all models, we control for propaganda fixed effects (δ_k) to compare reactions to the same propaganda release. Our Difference-in-Differences coefficient of interest is β_3, which shows the change in pro-ISIS rhetoric within one day of the propaganda release among users who were exposed to the propaganda.

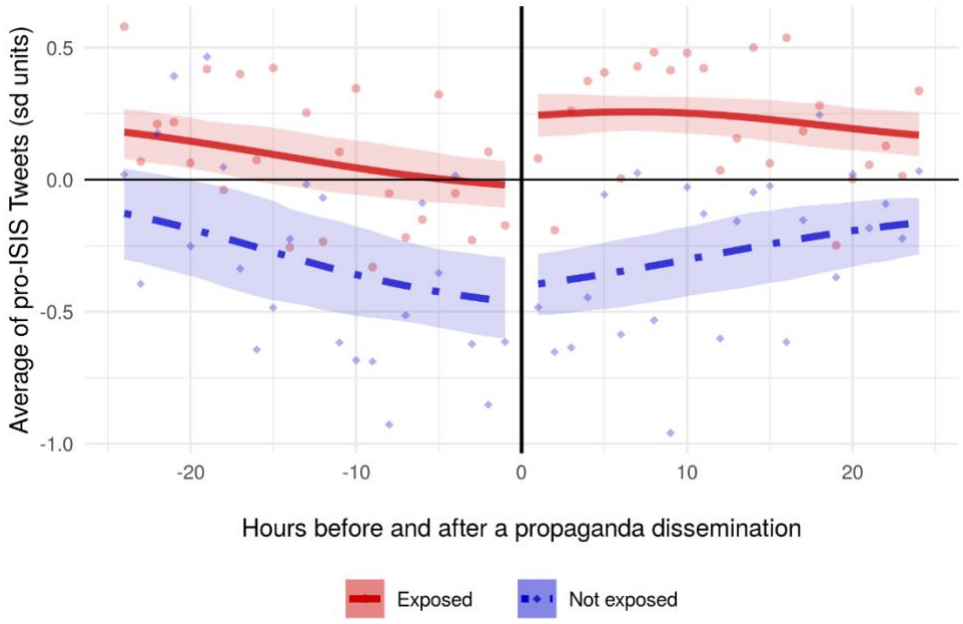


Figure 7: Pro-ISIS rhetoric: Parallel Trends. The Figure presents pre and post time trends in pro-ISIS rhetoric for the group of individuals exposed to ISIS propaganda (red) and the group of those who do not (blue). The x-axis is the number of days between the date of ISIS propaganda dissemination and the day in which ISIS followers posted on Twitter. The figure shows the average pro-ISIS content by each of the two groups in each day.

A potential concern with this research design is that individuals in the non-exposed group may have been exposed to propaganda through some other venue. For example, portions of a violent video may have been shown on television or shared in other online platforms, like Facebook. While we are not able to systematically rule out this kind of contamination, we are able to rely on two features of our data to strengthen the validity of our inference: (i) the high-frequency of our tweet-level observations, and (ii) the expected effects of contamination on our estimates.

The high frequency of Twitter posts allows us to measure instantaneous reactions to propaganda among a large sample of users. Even though some may have been exposed via other media, the fact that we observe changes on Twitter immediately after propaganda files were shared on the platform strengthens our confidence that the reactions are linked to these particular disseminations. Furthermore, the expected effect of contamination between exposed and non-exposed users should systematically bias our results to zero. If people in our control group are exposed to the propaganda in another media, we should expect to see their expressed sympathy for the ISIS mimic that of the treatment group, limiting our ability to detect a significant effect.[20] As we show in the next section, however, Twitter users who followed propaganda disseminators significantly changed their rhetoric after propaganda releases, while users who did not follow disseminators continued with the same rhetoric.

4.6 Results

We present our results in visual form, but report all tabular results in the Supplementary Materials. Figure 8 presents the Difference-in-Differences coefficients from model (1), where the top panel shows the results when the dependent variable is the pro-ISIS index, and the bottom panel breaks down pro-ISIS Twitter content by topic. When considering all types of ISIS

20 In fact, this is exactly what we find when examining the patterns before and after Twitter introduced the algorithmic timeline in February 2016, which moved away from showing the most recent tweets, to tweets that users would most 'like' to see. The new algorithm frequently showed users posts that their friends engaged with, which, in the context of our study, could lead to greater contamination of the results. In a test reported in the Supplementary Materials we find that before the switch there were strong differences in the reactions of exposed and non-exposed users, but after the switch, the reactions were not very different.

propaganda without taking into account its content (see leftmost points in Figure 8), we find that propaganda disseminations moderately increased support for the Islamic State among exposed followers. The change reflects about 2.5 percent increase over the baseline (pre-exposure) level of pro-ISIS rhetoric among exposed users. The bottom panel shows that this increase is driven by a higher number of posts sympathizing with ISIS ('ISIS sympathy') and content describing travel to Syria or foreign fighters ('Syria travel/FF').

Next, we examine how specific recruitment messages conveyed in each release impacted the group's online followers. Figure 3 shows ISIS followers' reactions to propaganda relating to the recruitment themes discussed above: (1) grievances; (2) ideological indoctrination; (3) material benefits; (4) social benefits; and (5) militancy and violence.[21] Consistent with the theoretical mechanisms, we find that many recruitment messages generated support for ISIS on Twitter. For example, propaganda appealing to *individual or group-based grievances*, such as content depicting Western countries' atrocities against Muslims or anti-Muslim hostility in the West, increased pro-ISIS content by about 11 percent. This increase is driven mainly by expressions of sympathy with ISIS and discourse on foreign fighters in ISIS-controlled territories (bottom panel of Figure 8). We also find that propaganda conveying *ideological indoctrination* – including messages calling for jihad, content conveying motivational speeches, and messages showing various religious activities – led to significant increases in pro-ISIS discourse on Twitter.

21 In these estimations, each propaganda release was coded for the inclusion of content on these themes if the majority of the relevant content features (described in Table 1) were detected in it.

In addition, we find that propaganda conveying *material inducements* and content emphasizing *social benefits and a sense of belonging* led to large increases in pro-ISIS posts. Messages showing the provision of public goods in ISIS territories and the benefits of marriage and family life increased support for the Islamic State on Twitter by about 9 percent, and propaganda displaying comradery between ISIS soldiers increased pro-ISIS content by over 13 percent. Given that these responses were measured within one day of exposure, the changes are substantively meaningful. Prior literature has pointed to the power of friendships and social networks, as well as the provision of public goods by armed groups, as important drivers of recruitment (Berman & Laitin, 2008; Wiktorowicz, 2005; Mousseau, 2011). Our findings show that these mechanisms were present in ISIS's internet recruitment efforts – and that they were successful in generating sympathy for the organization.

However, in contrast to the popular view that violent content was appealing to those who followed the Islamic State on the internet, we find that propaganda containing militancy and violence did not increase support for the group. The rightmost points in figure 8 show that within the first day of the release of violent propaganda, ISIS followers significantly *decreased* expressions of support for the group. After exposure to violent scenes, ISIS followers on Twitter were less likely to express sympathy with the group, discuss life in its controlled territories or mention foreign fighters. Instead, they talked more about events in the Syrian civil war, as can be seen in the bottom panel of Figure 8.

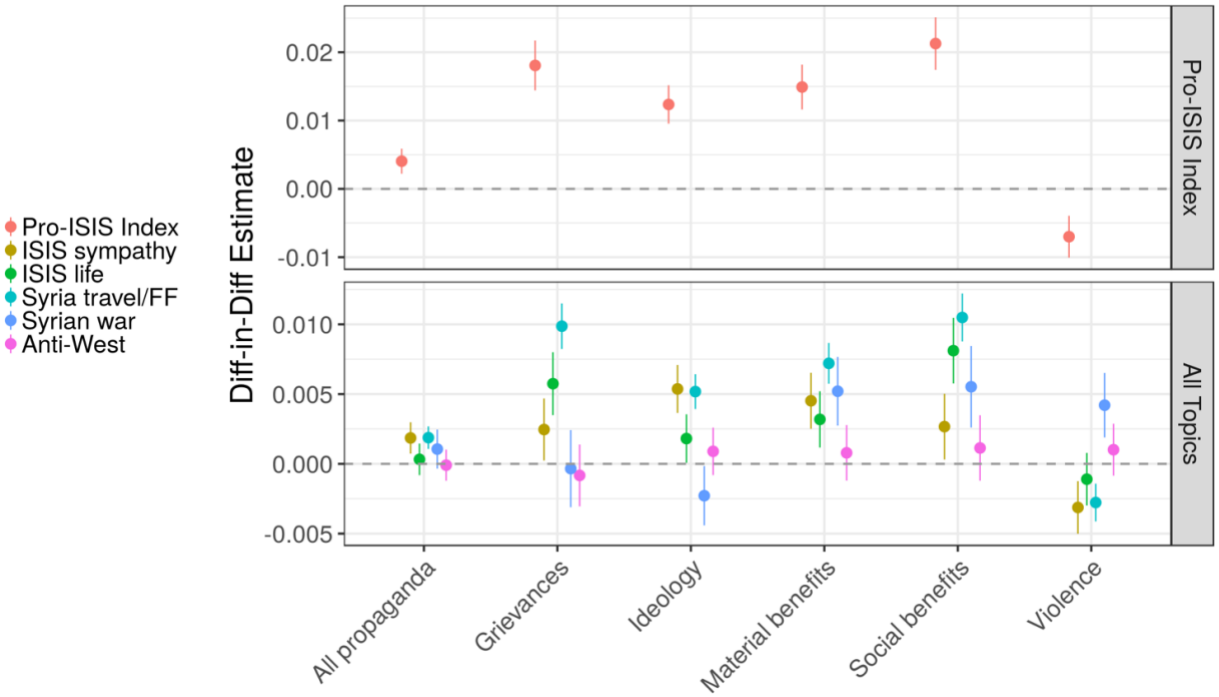


Figure 8: Propaganda and Pro-ISIS Rhetoric on Twitter. The figure presents coefficients estimated from a Difference-in-Differences analysis of the relationship between propaganda disseminations and pro-ISIS rhetoric on Twitter, captured within a one day before and after propaganda disseminations.

4.6.1 The Impact of Violent Propaganda

To delve a bit deeper into the specific types of violent content that were driving these findings, we report in Panel (A) of Figure 9 results for different types of violent imagery: scenes with air and artillery fire, content showing military and soldiers, scenes showing ruins after explosions, as well as content displaying violence against humans. We find that while content showing air and artillery and scenes with military and soldiers increased support for ISIS, propaganda showing brutal violence dramatically decreased positive commentary on the group.

Since ISIS propaganda videos sometimes included violent imagery in conjunction with non-violent themes, we show in Panel (B) of Figure 9 followers' reactions to non-violent messages when violent imagery was also present in the videos. In the figure, purple circles show the reactions

to propaganda videos containing little or no violent scenes, and black triangles show the reaction to videos with many violent scenes. The figure shows a very clear pattern: propaganda videos speaking to grievances, ideology, material benefits, and the social benefits of joining ISIS led followers to increase expressions of support for the group when the propaganda contained little or no violent imagery. However, when the same type of content appeared alongside violent scenes, ISIS followers on Twitter significantly decreased expressions of support for the group.

Did exposure to violent videos shape subsequent reactions to propaganda? To answer this question, we created a measure that counts, for each user and each propaganda release, the number of videos showing brutal violence to which they were exposed in the preceding weeks and months.22 Panel (C) in Figure 9 presents the distribution of this variable, showing that exposure to violent propaganda was not very frequent for the majority of users in our dataset; only a small proportion of users was exposed to more than ten videos displaying brutal violence over 2015 and 2016.

In Panel (D) of Figure 9, we examine how ISIS followers reacted to different types of propaganda messages after they have already been exposed to violent propaganda in prior videos. We created a binary variable that is coded 1 for users who had high levels of prior exposure to violent content, and 0 otherwise.[23] We use this variable to examine whether negative reactions to

22 That is, we counted the number of violent propaganda videos that each user was exposed to before each new exposure. This measure is cumulative over the two years of data in our study.

23 We define high levels of prior exposure to violence to be when the number of violent videos exposed in previous time periods is above the mean.

violent propaganda persist over time—that is, whether they can still be observed in user reactions to future videos, either violent or not.

The results suggest that violent propaganda did, in fact, have persisting effects. Videos containing non-violent themes did not generate positive reactions when users had prior exposure to ISIS's brutalities. As can be seen in Panel (D), the Diff-in-Diff coefficients for propaganda videos speaking to grievances, ideology, and the material and social benefits of joining ISIS are overall negative and statistically insignificant for users who were exposed to violent propaganda in previous time periods. This is important, as it sharply contrasts with our findings in Figure 5 that shows that non-violent themes in ISIS propaganda increased support for the group.

These results thus point to previously unexplored effects of extremist online propaganda. Even though the Islamic State became famous for its brutality, we find that what generated online support for the group was not its brutal violence, but the more 'positive' messages that conveyed the material, social, and spiritual benefits of recruitment.

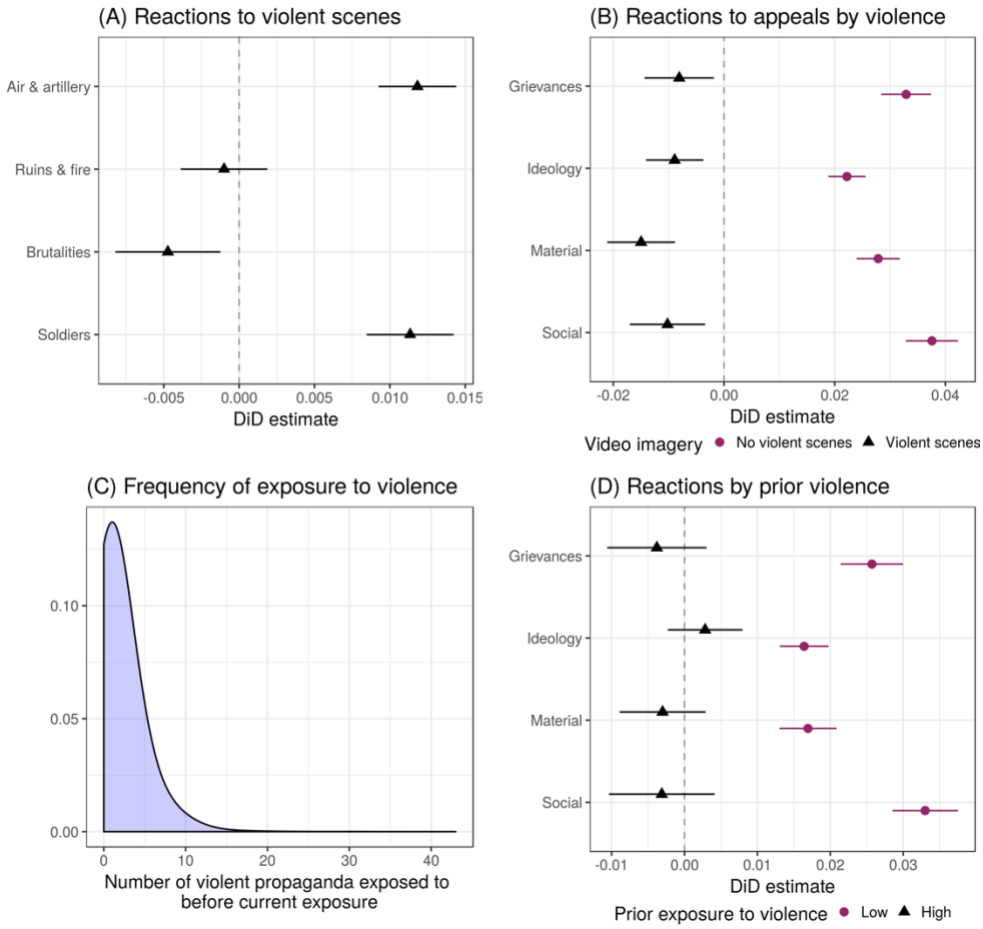


Figure 9: Reactions to Violent Propaganda. Panel (A) shows changes in pro-ISIS rhetoric on Twitter after exposure to different kinds of violent imagery in ISIS propaganda videos. Panel (B) shows how reactions varied when violent scenes appeared in the videos, and when they did not. Panel (C) shows the distribution of a variable that counts, for each user and each propaganda release, the number of videos showing brutalities to which each user was exposed in the preceding weeks and months. Panel (D) shows how reactions to propaganda varied between users who had low and high prior exposure to violent propaganda.

4.6.1.1 Why Produce Violence?

Why would the Islamic State continue to produce violent videos – and produce them in great numbers – if violence decreased online support among its followers? One theory is that the group was using violent videos as a selection mechanism to identify and recruit the more zealous

followers – the ones more likely to join the group and engage in violence – even if it alienated more moderate followers.[24]

In order to determine if this is true, we divided ISIS followers in our dataset into different groups, on the basis of their level of expressed sympathy with the group in the week prior to propaganda exposure. The variation in pro-ISIS rhetoric is quite large, with those most supportive of the group posting more than three times as many pro-ISIS tweets in the week preceding propaganda releases than the rest of the users in our sample.[25] We refer to those who expressed high levels of support for ISIS as more 'extreme' than those who did not, and examine whether their reactions to brutal violence were different from the rest of the followers.

Panel (A) of Figure 5 shows that extreme followers—those above the 75th percentile of pre-exposure support for ISIS—significantly increased positive tweets about ISIS after violent propaganda appeared on their timelines, while more moderate followers decreased their support for the group after exposure to violence. Panel (B) presents how extreme users' reactions to violence varied with prior exposure to brutal violence. If radicalized individuals become desensitized to brutal violence, we would expect their reactions to violent content to be more positive, on average, after they have already been exposed to a lot of violent propaganda. Indeed, this is what we find in Panel (B), which shows that extreme users' reactions to violent propaganda was even more positive among users who already had high levels of exposure to violence in prior propaganda videos. While not addressed within our theoretical framework explicitly, this pattern

24 Walter, The Extremist's Advantage, 2018.

25 Specifically, those in the top 10% posted a total of 1,908,297 pro-ISIS tweets in the week prior to propaganda releases in our sample, while all the rest posted a total of 551,625 pro-ISIS tweets.

corroborates the intuition underlying many models of radicalization—that individuals who move further down the path of extremism become more tolerant to (or even supportive of) brutal violence.

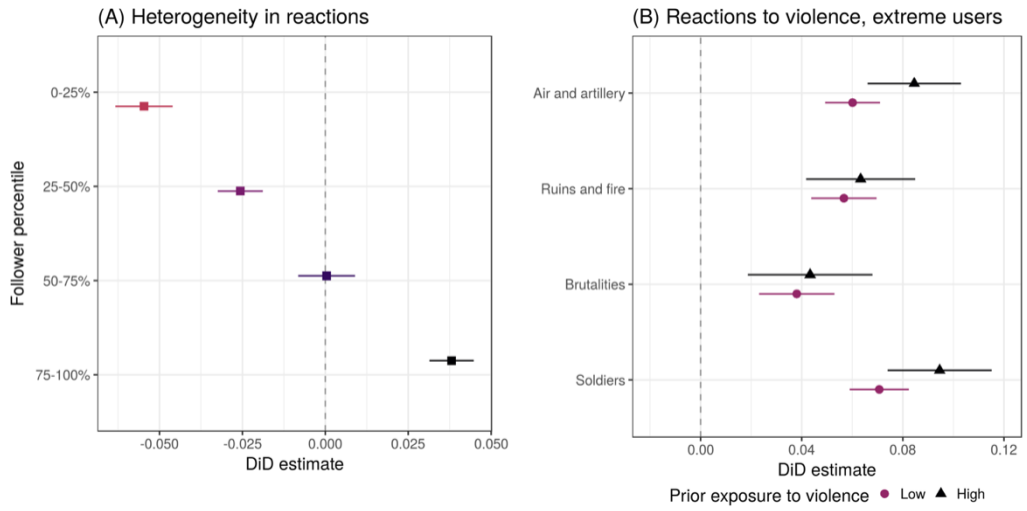


Figure 10: **Reactions to Violent Propaganda by Follower Type.** Panel (A) shows that ISIS followers who expressed high (low) levels of support for ISIS in the week prior to propaganda disseminations increased (decreased) their support for ISIS after exposure to violent videos. Panel (B) shows that extreme users' positive reactions to violent content were even stronger among those who had high levels of prior exposure to violent propaganda.

4.7 Conclusion

What does this research tell us about the type of counter-messages that are likely to work? We believe that incorporating knowledge of the messages that encourage support for groups like ISIS can improve existing efforts to combat online extremism. Knowing what types of propaganda are most effective—and least effective—provides a strong foundation for developing better counter-extremism strategies. As there are already many initiatives, and infrastructure in place, to

combat online propaganda and hate speech,[26] devising interventions based on evidence such as ours will constitute an important step forward.

As the past decade as shown us, internet platforms offer violent extremist groups cheap and effective ways to reach a global audiences in an unprecedented scale. They provide means to simultaneously increase breadth (radicalize moderate supporters) and depth (further radicalize hardline supporters) while disseminating content at a large scale. The method we developed for this study illustrates how powerful machine learning techniques can be applied to detect granular messages in large amounts of online materials, identify their dissemination on social media platforms, and measure the reactions of those exposed them in real time. By applying this method to Islamic State propaganda, we provided, to our knowledge, the first large-scale empirical study showing how different messages in ISIS propaganda impacted the group's sympathizers on Twitter.

We believe that this method can be easily replicated to study other online recruitment campaigns. Experts on radicalization have long argued that the process of radicalization is very similar across ideologies, whether the person is a jihadist, a white supremacist, or something else. Far-right extremists, who have considerably increased their presence on social media over the past decade, are employing many of the same substantive appeals as Islamist extremists, even though

26 For example, see "The Redirect Method: A Blueprint for Bypassing Extremism." at https://redirectmethod. org/ and "Counter Conversations: A Model for Direct Engagement with Individuals Showing Signs of Radicalisation Online" at http://www.isdglobal.org/wp-content/uploads/2018/03/Counter-Conversations_ FINAL.pdf

they target very different audiences.[27] Using our method to study the impact of white supremacist propaganda, for example, could break ground in understanding the appeal of these groups as well.

The results of this study can also inform content moderation strategies by social media companies. Technology platforms such as Google, Facebook and Twitter, have been dedicating significant resources to automate the detection and take down of violent content from their websites. Facebook, for example, massively expanded its personnel working on 'dangerous organizations' – an expert team that focuses on designing policies and building algorithms to block violent materials before they are uploaded (Bickert & Fishman, 2017; Alba et al., 2019). Google designed tools that redirect searches for violent propaganda towards alternative content challenging extremist worldviews, while removing millions of violent videos from YouTube. Twitter has stepped up its efforts to suspend accounts that disseminate violence, and since 2017, a large number of tech companies have been sharing 'digital fingerprints' to coordinate blocking of terrorist content across platforms—90% of which consist of graphic violence or content glorifying violent acts.[28] While there is no question that it is important to take down violent content, our results suggest that it will be fruitful to also focus on other, perhaps more subtle and less violent

27 Alice Marwick and Beca Lewis, "The Online Radicalization We're Not Talking About," New York Magazine, May 18, 2017.

28 For example, members of the Global Internet Forum to Counter Terrorism, which includes companies like Facebook, Twitter, Google, and Microsoft, have been coordinating the take down of terrorist propaganda since 2017. These companies have been sharing over 200,000 digital fingerprints of terrorist images and videos, 90% of which consist of graphic violence or content glorifying terrorist acts. See https://gifct.org/transparency/.

types of online messages, which, as our study shows, can be very appealing to followers of extremist groups.

Finally, even though our study focused on campaigns by extremist groups, our method is highly relevant to the study of other types of online campaigns. This includes campaigns surrounding democratic elections, disinformation campaigns designed to bolster autocrats, and intimidation campaigns aimed at suppressing protests and dissent. Our study's methodology reveals how social scientists can begin to study these important emerging phenomena. As many aspects of social and political life are influenced by exposure to online media, these methods are only going to become more relevant.

4.8. Acknowledgements

Chapter 4, Propaganda and Radicalization in an Internet Age, is co-authored. This paper is the product of a collaboration between Tamar Mitts, Barbara Walter, and me that began when Tamar approached us about the possibility of matching her data on social media radicalization and our data on online media produced and disseminated by the Islamic State (that I had begun collecting as a graduate student researcher with Prof. Barbara Walter). This paper then became the most extraordinary – and extraordinarily rewarding – collaborative venture of my time in grad school. To say working with Prof. Mitts and Prof. Walter is life-changing would be an egregious understatement. I will forever be grateful for their mentorship and guidance during our time working on this paper together.

Given that this piece is co-authored, it is perhaps most important for me to establish my contributions to the paper. I contributed to this paper in four ways:

5. I collected the online media data used to track disseminations of Islamic State propaganda on Twitter. This was primarily done as a GSR while working with Prof. Walter.
6. I contributed to our understanding of the theoretical drivers of extremist online engagement that serve as the exposition and background to the paper. Over the course of several drafts of the paper, I contributed significantly to our working understanding of the appeals used in propaganda from the perspective of the literature on radicalization and extremism in civil war.
7. I contributed significantly to the empirical design of this paper – most actively in the brainstorming and study design phase. In the early stages of this paper, Prof. Mitts, Prof. Walter and I were at a crossroads in determining how we both classify media releases into meaningful content categories and empirically show the effect of each content category on expressed preferences and behavior on social media. I brought the idea of reducing the words spoken in the video to transcriptions, while Tamar added that we could use object recognition to identify the objects that appear to enhance our detection of potential violence. This piece of the empirical design then served as the foundation for our strategy of disseminating tracing, which tracked disseminations of media scoring high on thematic content categories and traced the changes in online behavior across groups of individuals matched on pre-exposure expressions of support for the Islamic State who were exposed to media and those who were not.
8. I piloted early components of our empirical strategy, including verifying the usability and implementation of Amazon Rekognition in detecting objects that could indicate the presence of violent content in media. While many of the downstream empirical tasks were carried out by Professor Mitts and her graduate student researchers, I played a key role in

piloting these methods before we finalized and proceeded to executing our empirical strategy.

Outside of these contributions, I acknowledge the extraordinary efforts carried out by Profs. Walter and Mitts that made this paper possible.

4.8 References

Alba, D., Edmondson, C., & Isaac, M. (2022). *Facebook Expands Definition of Terrorist Organizations to Limit Extremism (Published 2019)*. Nytimes.com. Retrieved 14 March 2022, from https://www.nytimes.com/2019/09/17/technology/facebook-hate-speech-extremism.html.

Bass, L. (2014). *What Motivates European Youth to Join ISIS?*. Syria Comment. Retrieved 14 March 2022, from https://www.joshualandis.com/blog/push-factors-helping-isil-recruitment-loretta-bass/.

Baugut, P., & Neumann, K. (2019). Online propaganda use during Islamist radicalization. *Information, Communication &Amp; Society*, *23*(11), 1570-1592. https://doi.org/10.1080/1369118x.2019.1594333

Bickert, M., & Fishman, B. (2017). Hard questions: How we counter terrorism. *Facebook Newsroom*, *15*(06), 2017. Retrieved 14 March 2022, from https://about.fb.com/news/2017/06/how-we-counter-terrorism/

Birnbaum, E. (2019). Social media giants restructure counterterrorism effort into independent group with staff. The Hill. Retrieved 14 March 2022, from *https://thehill.com/policy/technology/462691-social-media-giants-restructure-counter-terrorism-effort-into-independent*

Blazak, R. (2001). White Boys to Terrorist Men. *American Behavioral Scientist*, *44*(6), 982-1000. https://doi.org/10.1177/00027640121956629

Borum, R. (2011). Radicalization into Violent Extremism I: A Review of Social Science Theories. *Journal Of Strategic Security*, *4*(4), 7-36. https://doi.org/10.5038/1944-0472.4.4.1

Bouchard, M. (2015). *Social networks, terrorism and counter-terrorism: Radical and connected*. Routledge.

Carter, J. A., Maher, S., & Neumann, P. R. (2014). Greenbirds: Measuring importance and influence in Syrian foreign fighter networks. The International Centre for the Study of Radicalisation and Political Violence. *King's College, Londra*.

Crenshaw, M. (2011). *Explaining Terrorism: Causes, Processes, and Consequences*. New York: Routledge.

Dalgaard-Nielsen, A. (2010). "Violent radicalization in Europe: What we know and what we do not know." *Studies in Conflict & Terrorism* 33(9):797–814. https://doi.org/10.1080/1057610X.2010.501423.

Dandachli, R. (2017). Fighting ideology with ideology: Islamism and the challenge of ISIS. Brookings. Retrieved 14 March 2022, from *https://www.brookings.edu/opinions/fighting-ideology-with-ideology-islamism-and-the- challenge-of-isis/*

Della Porta, D. (2018). Radicalization: A Relational Perspective. *Annual Review of Political Science* 21:461–474. https://doi.org/10.1146/annurev-polisci-042716- 102314.

Friedman, J., Trevor H., & Robert T. (2010). Regularization Paths for General- ized Linear Models via Coordinate Descent. *Journal of Statistical Software* 33(1):1–22. https://doi.org/10.18637/JSS.V033.I01.

Gómez, Á., López-Rodríguez, L., Sheikh, H., Ginges, J., Wilson, L., & Waziri, H. et al. (2017). The devoted actor's will to fight and the spiritual dimension of human conflict. *Nature Human Behaviour*, *1*(9), 673-679. https://doi.org/10.1038/s41562-017-0193-3

Greenberg, Andy. (2016). Google's Clever Plan to Stop Aspiring ISIS Recruits. Wired. Retrieved 14 March 2022, from *https://www.wired.com/2016/09/googles-clever-plan-stop-aspiring-isis-recruits/*

Horgan, J. (2008). From profiles to pathways and roots to routes: Perspectives from psychology on radicalization into terrorism. *The ANNALS of the American Academy of Political and Social Science* 618(1):80–94. https://doi.org/10.1177/0002716208317539.

Horgan, J., Taylor, M., Bloom, M., & Winter, C. (2016). From Cubs to Lions: A Six Stage Model of Child Socialization into the Islamic State. *Studies In Conflict &Amp; Terrorism*, *40*(7), 645-664. https://doi.org/10.1080/1057610x.2016.1221252

Isaac, M. (2016). Twitter Steps Up Efforts to Thwart Terrorists' Tweets. New York Times. Retrieved 14 March 2022, from *http://www.nytimes.com/2016/02/06/technology/twitter-account-suspensions- terrorism.html*

Jackson, B. (2019). *Practical Terrorism Prevention: Reexamining U.S. National Approaches to Addressing the Threat of Ideologically Motivated Violence* (1st ed., pp. 1-335). Rand Corporation.

King, G., & Langche Z. (2001). Logistic regression in rare events data. *Political analysis* 9(2):137–163. https://doi.org/10.1093/oxfordjournals.pan.a004868

Kydd, A. H., & Barbara F. W. (2006). The strategies of terrorism. *International Security* 31(1):49–80. https://doi.org/10.1162/isec.2006.31.1.49.

Lyons-Padilla, S., Gelfand, M., Mirahmadi, H., Farooq, M., & van Egmond, M. (2015). Belonging nowhere: Marginalization & radicalization risk among Muslim immigrants. *Behavioral Science &Amp; Policy*, *1*(2), 1-12. https://doi.org/10.1353/bsp.2015.0019

Malley, R., & Jon F. (2018). The Long Shadow of 9/11: How Counterterrorism Warps US Foreign Policy. *Foreign Aff.* 97:58.

McCauley, C., & Sophia M. 2008. Mechanisms of political radicalization: Path- ways toward terrorism. *Terrorism and political violence* 20(3):415–433. https://doi.org/10.1080/09546550802073367.

Mitts, T. (2019). From isolation to radicalization: anti-Muslim hostility and support for ISIS in the West. *American Political Science Review* 113(1):173–194. https://doi.org/10.1017/S0003055418000618.

Mousseau, M. (2011). Urban Poverty and Support for Islamist Terror: Survey Results of Muslims in Fourteen Countries. *Journal of Peace Research* 48(1):35–47. https://doi.org/10.1177/0022343310391724.

Nagle, A. (2017). The Lost Boys: The Young Men of the Alt-Right Could Define American Politics for a Generation. The Atlantic. Retrieved 14 March 2022, from *https://www.theatlantic.com/magazine/archive/2017/12/brotherhood-of-losers/544158/*

Neumann, R. (2013). The trouble with radicalization. *International Affairs* 89(4):873–893. https://doi.org/10.1111/1468-2346.12049.

Revkin, R. (2019). Relative legitimacy and displacement decisions during Rebel Governance. 1-36. *http://dx.doi.org/10.2139/ssrn.3365503.*

Rink, A., & Sharma, K. (2016). The Determinants of Religious Radicalization. *Journal Of Conflict Resolution, 62*(6), 1229-1261. https://doi.org/10.1177/0022002716678986

Rosen, G. (2019). A Further Update on New Zealand Terrorist Attack. *Facebook Newsroom*. Retrieved 14 March 2022, from *https://newsroom.fb.com/news/2019/03/technical-update-on-new-zealand/*

Sageman, M. (2004). *Understanding terror networks*. University of Pennsylvania Press. Sedgwick, Mark. 2010. "The concept of radicalization as a source of confusion." *Terrorism and Political Violence* 22(4):479–494.

Thrall, A. T., & Goepner, E. (2017). Step back: lessons for US foreign policy from the failed war on terror. *Cato Institute, Policy Analysis*, (814).

Vidino, L., & Hughes, S. (2015). Countering violent extremism in America. *The George Washington University Center for Cyber and Homeland Security*. Retrieved 14 March 2022, from https://extremism.gwu.edu/sites/g/files/zaxdzs2191/f/downloads/CVE%20in%20America.pdf

Walter, B. (2017). The Extremist's Advantage in Civil Wars. *International Security*, *42*(2), 7-39. https://doi.org/10.1162/isec_a_00292

Weimann, G. (2016). Terrorist migration to the dark web. *Perspectives on Terrorism* 10(3):40–44. Retrieved 14 March 2022, from https://www.jstor.org/stable/26297596

Wiktorowicz, Q. (2005). *Radical Islam rising: Muslim extremism in the West*. Rowman & Littlefield Publishers.

Wilner, A., & Dubouloz, C. (2010). Homegrown terrorism and transformative learning: an interdisciplinary approach to understanding radicalization. *Global Change, Peace &Amp; Security*, *22*(1), 33-51. https://doi.org/10.1080/14781150903487956

Wilner, Alex S and Claire-Jehanne Dubouloz. 2010. "Homegrown terrorism and transformative learning: an interdisciplinary approach to understanding radicalization." *Global Change, Peace & Security* 22(1):33–51.

APPENDICES

Definitional Appendix

Defining Extremist Groups

Before continuing, it is important to establish what constitutes an extremist group – a definition that has been the cause of debate among scholars across disciplines. At its root, the definition of extremism as a political construct identifies an ideology far removed from the local political, religious, social, or cultural mainstream (Coleman & Bartoli, 2009). However, extremism in this sense can extend beyond the ideological beliefs of individuals into the methods through which extremists, as political actors, seek to achieve their aims. Extremist methods often showcase a disregard for liberties, lives, and humanity that actors are willing to display in order to accomplish their goals (Neumann, 2010). The range of ideologies, behaviors, goals, and beliefs that can fall under the umbrella of extremism is vast – as are the methods that extremists can employ to accomplish their goals.

Extremist beliefs develop and spread through radicalization – a process of persuasion through which people develop extremist beliefs or adopt those of an existing organization (Borum, 2011). Radicalization, as a process, is broad – it does not, in and of itself, imply that individuals support violent behavior, or seek to participate in violence themselves. Rather, it simply means that individuals have adopted beliefs, goals, behaviors, or methods that differ from the locally accepted norms. Furthermore, radicalization as a concept does not imply that individuals were directly exposed to another extremist ideology. Individuals may self-radicalize in the face of social, political, cultural, or religious adversity or oppression without the presence (Trip et al., 2019; Horgan, 2009). When other individuals are involved in the radicalizing process, rather, I refer to this concept more specifically as *indoctrination* to separate it from the broader phenomenon.

It is possible to hold extremist ideological beliefs and political goals without subscribing to methods of achieving those goals through violence. This points to a core tension between the academic and policy-maker perspectives on extremism: while the range of possible extremist groups and the methods through which they seek to achieve their goals is vast, the political label of extremism, and its legal consequences, are typically reserved for violent extremists that engage in acts of terrorism. In the UK, scholars have noted that legal definitions that single out groups for specific ideological beliefs linked to the overt support of violence are subject to manipulation by organizations that hold extremist beliefs, but shift their strategies for accomplishing their goals into the realm of mainstream politics (Allen, 2019). In the United States and Europe alike, an extensive focus on violent Islamic extremism and the tactics of terror pursued by some of these groups has narrowed public discourse on which extremist groups pose a threat to public safety almost solely to these organizations and their followers, while focusing attention almost entirely on the strategy of terrorism that such groups may employ (Allen, 2019).

By breaking down three strategies by which extremists attract support and secure compliance, my arguments in this chapter suggest that the limitations imposed on public understandings of extremism as defined by political leaders harm our efforts to effectively identify and counter extremist influence. By pigeon-holing extremism to a particular set of political objectives and strategies, many contemporary policies have made it easier for extremists to change the way individuals view them by changing the way that they market their ideology, goals, or methods (Allen, 2019; Borum, 2011). Groups that hold politically extreme goals and who are willing to resort to extreme violence to accomplish them can reframe their appeals to individuals not sympathetic to their cause – linking their political struggle to other, more mainstream, struggles that interest more moderate constituencies and emphasize the material benefits of supporting their

political rise. Furthermore, groups committed to extreme goals that go on to turn away from strategies of terrorism are more easily able to shed labels of extremism tied to their terrorist acts under the law. This politicization of how different audiences conceptualize extremism highlights an opening that extremist groups can exploit when seeking support: if they can manipulate the way that individuals perceive the costs and benefits of what they do or what they believe, they may be able to co-opt a broader audience that doesn't subscribe to their ideology to join their cause.

To accommodate the strategic component of extremist messaging, this paper adopts a broad definition of extremism, defining extremist groups as those groups whose political goals, beliefs, or tactics fall outside the locally acceptable means of political discourse and action. This definition is malleable. However, most importantly, this definition does not fix a group's status as extremist to the strategy it employs on the battlefield. Rather, this definition recognizes that extremists possess a wider toolkit when seeking to accomplish their goals. As their local priorities and capabilities change, the way they market themselves or seek to accomplish their goals may shift while their ideological identity remains the same.

Data Coding Appendix

A. Coding Section

The following section comprises the instructions given to human coders. Three different human coders hand-coded a training set of 482 releases of Islamic State media. This training set comprised of 280 video releases, 101 audio releases, and 101 text releases.

Familiarizing Yourself with the Sample Text

For this task, we will begin by familiarizing you with the sample Islamic State media data found in your Project folder. I have saved this as a .csv file for your convenience; it will open in

either Numbers or Excel. I have had more luck with Arabic in Numbers, but you may have a different experience.

You will notice eight primary column headers:

doc_id	text	prop_key	title_e	ideo_indoc	gsp_coop	covi_cii	civi_ci	opvi_oppvi

- **Doc_id** is the document's unique ID number. It helps us locate the file in a larger collection of files. It also allows us to merge this dataset with any other dataset using the doc_id as the merging element.
- **Text** is the arabic text taken from the video.
- **Prop_key** is the key text used to extract the titles from the original document.
- **Title_e** is the english language title transliterated (sometimes badly) by the author of the clearinghouse.
- **Link** is the link to the internet location of this file. You will use this to view the banners for the releases. Please do not watch the content. Some of the actual media content is extremely graphic.
- **Ideo, GSP, Covi, Civi, and Opvi** are all indicators for types of content, and I will elaborate on these below.

Relevant Content Categories

Ideo is Ideological Indoctrination (Strategy of Indoctrination)

What it is

Ideological indoctrination in this project is the presentation of ideological material -- content forwarding a particular political, religious, social, or economic argument corresponding to

the stated group purpose -- as the justification for, guiding principle of, or otherwise central component corresponding to the production of a piece of media.

What it looks like in words and images

This content features individuals from rebel organizations providing speeches, sermons, or illustrative lessons regarding the description of, importance of, rewards attached to, or punishments for not following the ideology endorsed by the group. Group ideology may include religious beliefs and interpretation, political beliefs and interpretation of their implementation, and social policies and beliefs that follow from ideological commitments. This content may also include animations, illustrated companions, audio, or text descriptions of any of these listed or related activities.

How you should code it

If you observe the features described above, put a 1 in the Ideo column of the sheet. If you do not, leave that cell blank.

GSP is Goods and Services Provision (Strategy of Co-optation)

What it is

Goods and services provision in this project is the distribution of goods like food, water, heating and cooking oil, or other supplies and the provision of services like public entertainment spaces, road repair, electrical provision and repair, water services, and other functions of the state or community that are provided by an actor seeking to step in for, supplement, or supplant the state.

What it looks like in words and images

This content features individuals from the rebel organization providing goods and/or services to people within the territory it controls. Potential public goods include the distribution of food, potable water, or cooking gas. Potential public services include the distribution of health

services, education, currency distribution and exchange, and policing. In the Islamist context, at least some of this is carried out in the name of the distribution of زكاة benefits, and the distibution -- particularly the widespread distribution -- of زكاة benefits qualifies as GSP.

How you should code it

If you observe the features described above, put a 1 in the GSP column of the sheet. If you do not, leave that cell blank.

Covi is Coercive Violence (Strategy of Intimidation)

What is it

Coercive violence is the use of violence to accomplish one of two goals: encourage civilians to behave in a certain way (e.g., in accordance with the policies that rebels have enacted) or deter them from behaving in others (e.g., deter them from rebelling or resisting rebel authority in places they influence or control). In this context, coercive violence is meant to employ violence in a way that gets people to act the way that the group wants them to act.

What it looks like in words and images

This content features individuals from rebel organizations using, threatening to use, or implying that they will use violent means to punish people for not acting the way that a rebel group would like them to act. This may include the execution or violent punishment of civilians for violation of rebel policies or resistance to rebel control as well as speeches, sermons, or religious quotations promising violent punishment for civilian transgression. This may also include public displays of force, such as armed soldier patrols enforcing policies (in the Salafist tradition, "beard patrols" wherein authorities patrol to ensure that men have beards of a particular grooming and length would qualify). This may also include the execution of non-military "spies" -- those accused of being informants for opposition groups or hostile countries.

How you should code it

If you observe the features described above, put a 1 in the Covi column of the sheet. If you do not, leave that cell blank.

Civi is Civilian Violence (Strategy of Intimidation)

What is it

Civilian violence here means *indiscriminate* or *genocidal* civilian violence -- violence directed at civilians either without a stated or easily interpretable purpose, violence simply to inflict harm against a group of people, or violence aimed at eliminating an entire group, class, or nation of people. This differs from coercive violence because the goal is either not clear, the goal is actually to inflict violence against civilians, or the goal is to eliminate a group, class, or nation of people rather than punish people for or deter people from specific transgressions.

What it looks like in words and images

This content features individuals from rebel organizations using, threatening to use, or implying that they will use violent means against civilians without a stated purpose, or specifically for the purpose of punishing a group of people for their existence. This may include the execution or violent punishment of civilians for no stated purpose, the massacre of civilians for no stated purpose, or the killing or massacre of people in order to eliminate them as a group, class, or nation.

How you should code it

If you observe the features described above, put a 1 in the Civi column of the sheet. If you do not, leave that cell blank.

Opvi is Oppositional Violence (Strategy of Intimidation)

What is it

Oppositional violence means violence directed at military, armed force, or police resistance to rebel groups. This violence is militaristic in nature; rebels are using violence as a tool to gain military control of a particular territory or group of people. This differs from the other two categories above because this violence is directed toward opposition military personnel. The tactics might be the same -- rebels may use execution, brutal tactics, or deterrent public violence, but the target here is the important distinguishing feature.

What it looks like in words and images

This content features individuals from rebel organizations using or threatening to use violent means against opposition military personnel. This may include the execution of military officials, military activity against opposition militaries that features the use of violence directly against opposition military personnel, or the public torture or beating of opposition military officials.

How you should code it

If you observe the features described above, put a 1 in the Opvi column of the sheet. If you do not, leave that cell blank.

IMPORTANT DISCLAIMER:

If at ANY TIME while coding this content, you feel uncomfortable with what you are doing and need to stop, please stop immediately and let me know. I will completely understand. Nothing is more important than your health, safety, and well-being.