Foreign Influence Operations in the Cyber Age

A dissertation presented by

Josh A. Goldstein

to

The Department of Politics and International Relations

in partial fulfilment of the requirements
for the degree of
Doctorate of Philosophy
in the subject of
International Relations

University of Oxford
Oxfordshire, United Kingdom

September 2021

Word Count: 78,850

Abstract: After Russia's interference in the 2016 U.S. election, influence operations gained great prominence in the public consciousness. Yet, International Security Studies lags behind. Our theories of public opinion formation are primarily from a pre-cyber age, and we lack answers to fundamental questions about influence operations as a tool of statecraft. How can we conceptualize influence operations? Is pretending to be a member of another state to influence foreign citizens a viable strategy? Does plausible deniability afford states the benefits theorized by existing covert action literature? I interrogate these questions to provide a better understanding of the psychological mechanisms and potential effects of foreign influence operations, contributing to the academic study of secrecy and to illuminating the threat democratic societies now face.

I begin with a concept-building paper, which proposes and defends a definition of influence operations and explains their prevalence today. This prepares the ground for an emerging research field and motivates the investigation of the benefits of secrecy. In the second paper, leveraging cross-national experiments, I challenge the conventional top-down model of cue-taking in International Relations. I show that elite cues do not have a significantly stronger impact than public cues on citizens' attitudes towards a cyber-attack. This suggests that influence operations that attempt to influence mass attitudes by falsifying peer cues may allow foreign actors to benefit from secrecy for operational success. The third paper builds from the observation that politicians increasingly outsource disinformation work to private firms, but the assumption that they benefit from plausible deniability is untested. Results from an original experiment show that when governments are not implicated in an operation that originates in their country, they can avoid blowback and yet are no less effective at stirring paranoia or appearing omnipotent. I add the novel insight that the country of origin matters much more to Americans than the specifics of attribution.

Acknowledgements

This dissertation is the academic output of a life chapter. It was one of learning and exploration, of long stretches of study and lighter nights at the pub. I am indebted to many friends, colleagues, mentors, teammates, and community members for the last four years. I was surrounded by a group of thinkers and people who strengthened my work and my character.

Over my time in graduate school, my advisor, Janina Dill, has had the biggest impact on my intellectual development. Thank you, Janina, for going above and beyond the normal scope of supervisor duties. You supported me at every step where others might not have: my interest in policy, my pivoting topics from the MPhil to the DPhil, my tight timeline for completion. Moreover, in helping me become a more logical thinker, you gave me a great gift. I see problems and arguments differently because of you. You did not give me dinner but taught me to fish.

I benefited from many formal and informal mentors at Oxford. Thank you to Dominic Johnson for supervising my first year of the MPhil and inspiring my interest in the psychological aspects of international security. When Janina went on leave, Andrea Ruggeri welcomed me to the Brasenose College Workshop and offered feedback on multiple drafts and early-stage ideas. My dissertation is better because of Dominic and Andrea's involvement. I also benefited greatly from feedback from my Confirmation of Status assessors, Rob Johnson and Duncan Snidal, and from the input of colleagues, including Laura Courchesne, Brian Cryst, Jan Des Fleurs, Jeff Ding, Andrew Eggers, Emily Feldman, Eyck Freymann, Ross Gildea, Jess Gliserman, Todd Hall, Monica Kaminska, Spyros Kosmidis, Katie Mann, Soledad Artiz Prillaman, Serik Sagymbayev, David Shimer, and Giuseppe Spatafora.

I am grateful for funding from the Clarendon Scholarship, the Smith Richardson Foundation, St. Hilda's College, and the Department of Politics and International Relations. From the Department, Andrew Melling, Lana Fisher, and Dawn Tohill always supported my applications for grants and fellowships. Thank you also to the St. Hilda's College staff, especially my college advisors Karina Cendon Boveda and Hannah Smith.

During the final DPhil year, I was lucky to split time between my dissertation work and a predoctoral fellowship with the Stanford Internet Observatory (SIO). Thank you to Elena Cryst and Alex Stamos for welcoming me, and to the entire team for making the past year such a rewarding one. David Thiel was instrumental to all things tech, including many of the takedown investigations that I refer to in the pages to come. Matt Masterson and Riana Pfefferkorn made SIO a place I love to work. Sam Bradshaw, Renée DiResta, Shelby Grossman, and Carly Miller each offered valuable feedback on different parts of this project and shaped my thinking on influence operations. Thank you to Shelby in particular, who throughout the pandemic has served as equal parts co-author, mentor, boss, and friend. You taught me OSINT skills, encouraged me to present ideas in new venues, and have made me a better researcher.

At CISAC, I thank Rose Gottemoeller, Steve Pifer, Harold Trinkunas, and the entire Fellows Cohort. Herb Lin and the Social Science Research Group offered feedback on my first paper. In the final stretch of the dissertation, Lindsay Hundley was a most supportive friend and colleague—helping me with data visualization, serving as a sounding board, and giving me the

final push that I needed. Mike Tomz and Jessica Weeks also provided instructive feedback in the final hours. Thank you for your generosity of time and for making academia a team sport.

I was lucky to have had such strong teachers at Harvard, who inspired my interest in international security and encouraged me to pursue a graduate degree at Oxford. Thank you to Graham Allison, Nicholas Burns, David Gergen, Justin Gest, Greg Honan, Tyler Jost, Meghan O'Sullivan, Steve Rosen, and Adam Siegel. Graham, through your class, I found my passion for international security. In the years after, you taught me innumerable lessons about research and policy. Steve, much of my interest in secrecy, reflected in this project, began years ago in our undergraduate thesis meetings. During graduate school, you were a crucial advisor on project ideas and encouraged my transition from studying public opinion formation to influence operations specifically. Thank you to Tarun Chhabra for the opportunity to work with your team at Brookings and for all your guidance since. Finally, any time I had an academic quandary to navigate or a life decision to make, Justin Gest was only a call away.

I am also thankful to earlier teachers from both Solomon Schechter and Fieldston who influenced me. They include Janis Birt, Liz Fernandez, Bobby Greenwald, Lisa Levine, Shelby Stokes, and Grace Yun.

Many friends at Oxford made England home. Valerie Bradley and Joel Kariel doubled as dear friends and R saviors. Joel, along with Ailish Saker, helped me navigate life across two continents. KG Kaelin was my always-ready lunch buddy and didn't hesitate to pack and ship my stuff across the Atlantic. Alex Herkert, Jeff Sload, and Eyck Freymann made Hill View and Jericho our own; Tyler Pager did so through FaceTime. Salil Patel and Margaret Irving were generous hosts in London, and Emily Feldman never let me stop laughing. My Oxford basketball family, especially coach Mathias Rufino, Jerry Amokwandoh, Fergus Proctor, and Grigalius Taujanskas, always kept me hustling. I appreciate the role that so many others played too.

Emma, thank you for all of your support during this chapter. You helped me separate work from life and find my own definition of balance. You also endured far more of my rants about international security than any one person should have to bear. TC and Joey, you welcomed me into the trio you all hold so close and taught me more about life than you will ever know. "Onward and through it!"

Finally, and most of all, I thank my family. My brothers, Zach and Adam, are my best friends. You poke fun at my nerdiness, but always have my back. My parents, Tara and Wayne, are my biggest supporters. They embraced, with grace and good spirit, having a grown son in the thick of dissertation crunch time move in during a global pandemic. You cheered me on for the whole race, but practically ran the last few miles with me. *We* did it.

I am blessed to have four grandparents who instilled within me, from a young age, the value of education. Their lesson led to my passion. Pursuing a PhD at Oxford would not have been possible without the sacrifices each of them made. "From generation to generation…" I stand on the shoulders of giants. This dissertation is dedicated to the four of you—Bobbie and Marty, Merry and Richie.

July 2021

Table of Contents

Abstract 1
Acknowledgements 2

CHAPTER 1: PUBLIC OPINION AND INTERNATIONAL SECURITY 5

The Problem 5
The Contribution in Brief 7
The Global Nature of Cyber-Enabled Influence Operations 13
Public Opinion About Foreign Policy 17
The Effect of Public Opinion On Foreign Policy 25
Public Opinion and the Resilience of Democracy 29

CHAPTER 2: FOREIGN INFLUENCE OPERATIONS AND INTERNATIONAL SECURITY STUDIES .. 32

Introduction 33
Conceptualizing Influence Operations 38
Why Influence Operations are Under-Explored in International Security Studies Scholarship 56
Why Security Studies Should Consider Influence Operations 61
Contributions of Influence Operations to International Security Studies 78
Forging Ahead: A Research Agenda 85
Appendix 87

CHAPTER 3: FOLLOW THE EXPERTS OR FOLLOW THE CROWD? EXPERIMENTAL EVIDENCE ON PUBLIC ATTITUDES TOWARDS A CYBER-ATTACK 93

The "Bandwagon Effect" and the Puzzle of the Elite-Public Divide 94
Previous Research on Cue-Taking and Public Opinion about Foreign Policy 97
Summary of Hypotheses 107
Experimental Design 108
Results 111
Exploratory Analysis 116
Conclusion 119
Appendix: Questionnaires 122
Appendix: Experiment 1 126
Appendix: Experiment 2 138

CHAPTER 4: PLAUSIBLE DENIABILITY AND OUTSOURCED INFLUENCE OPERATIONS 146

Introduction 147
Secrecy and Foreign Covert Action 152
Theorizing Forms of Uncertainty 158
Applying Logics of Plausible and Implausible Deniability to Influence Operations 160
Experimental Design 168
Results 173
Conclusion 186
Appendix 189

CHAPTER 5: CONCLUSION 218

Contributions & Implications 218
Future Research 222

BIBLIOGRAPHY 228

Chapter 1: Public Opinion and International Security

The Problem

"Oh f---, how did we miss this?"

That's how Mark Zuckerberg reacted in December 2016, when he was first briefed about Russian meddling during the 2016 U.S. presidential campaign.[1] The scope and scale of social media influence operations targeting the 2016 U.S. presidential election caught the tech platforms, the cybersecurity community, and, as far as publicly available evidence suggests, the U.S. government, by surprise.[2] We now know that the Russian Internet Research Agency's activity reached 126 million people on Facebook and at least 1.4 million on Twitter,[3] all with fewer than 100 employees and a monthly budget of $1.25 million.[4]

On the one hand, this failure of foresight should not be surprising. Facebook and Twitter were only created in 2004 and 2006 respectively. Since social media companies are relatively new, so too are the ways in which they can be manipulated by political actors.

On the other hand, the weaponization of social media by political actors was foreseeable. As I detail in the pages to come, states have long waged influence operations, defined here as "covert, informational efforts intended to affect attitudes or opinions in a manner favorable to the perpetrator." From Spartan rumor campaigns smearing Athenian statesmen during the Peloponnesian War, to Benjamin Franklin's fake newspaper printings in the 1770s in France, to

[1] Note that Facebook's internal security team had found evidence of Russian meddling nine months earlier, but even a month after the election the findings had yet to make their way to Zuckerberg. Frenkel, Sheera and Cecilia Kang. 2021. *An Ugly Truth: Inside Facebook's Battle for Domination.* HarperCollins Publishers, as cited in Allen, Mike. "'An Ugly Truth': How Facebook discovered Russian meddling.'" *Axios*. July 9, 2021. Accessed: https://www.axios.com/facebook-book-russia-zuckerberg-sandberg-3bcb04fb-b8e9-4e8e-83d1-9a00e1218a27.html

[2] This is not to suggest that the U.S. government did not know about interference prior to election day. Public reporting suggests it did. Rather, these sectors had not adequately anticipated the scale and spread of social media influence operations—as they would come to occur in the United States and across the globe.

[3] DiResta, Renée, Kris Shaffer, Becky Ruppel, David Sullivan, and Robert Matney. 2019. "The Tactics & Tropes of the Internet Research Agency." *New Knowledge Report*. p. 6. Accessed: https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1003&context=senatedocs

[4] Madrigal, Alexis C. "Russia's Troll Operation Was Not That Sophisticated." *The Atlantic*. February 19, 2018. Accessed: https://www.theatlantic.com/technology/archive/2018/02/the-russian-conspiracy-to-commit-audience-development/553685/

"black" radio stations used by both Axis and Allied powers during World War II, influence operations have long been used against allies and adversaries, in both peacetime and war.[5]

If influence operations are a regular staple of international affairs, why has their recent wave taken policymakers, and, as I will argue below, political scientists by surprise? The illustrative examples suggest that the tools used by state actors for influence operations are highly dependent on the technology of the day. By the mid-2010s, social media was a sprawling communication channel, one that offered newfound speed, scale, and anonymity. Leveraging fake social media accounts and pages, state actors could more easily target foreign publics than in the past while foregoing the risks of placing agents-of-influence into foreign jurisdiction.

In the years after 2016, a budding community of defenders coalesced—including teams at social media platforms dedicated to the influence operation challenge, private companies with proprietary techniques to identify coordinated behavior, and even university centers that seek to understand how information moves online.[6] From different pockets of academia, studies have examined the tactics and types of accounts run by the Internet Research Agency on Twitter,[7] how Internet Research Agency accounts were cited by real news outlets,[8] the types of social media

---

[5] On the Peloponnesian War: Pearson, Lionel. 1936. "Propaganda in the Archidamian War." *Classical Philology* 31(1) 33–52; On U.S. propaganda in France: Waller, J. Michael. 2009. "The American Way of Propaganda: Learning from the Founding Fathers." *Strategic Influence: Public Diplomacy, Counterpropaganda, and Political Warfare* Washington, D.C.: The Institute of World Politics Press.; Sulick, Michael J. 2012. *Spying in America: Espionage from the Revolutionary War to the Dawn of the Cold War*. Georgetown University Press.

[6] The Carnegie Endowment for International Peace's Partnership for Countering Influence Operations (PCIO) created a dataset of 460 counter-influence operations initiatives. See: Smith, Victoria. 2020. "Mapping Worldwide Initiatives to Counter Influence Operations." *Carnegie Endowment for International Peace*. Accessed: https://carnegieendowment.org/2020/12/14/mapping-worldwide-initiatives-to-counter-influence-operations-pub-83435

[7] Bastos, Marco and Johan Farkas. 2019. "'Donald Trump Is My President!': The Internet Research Agency Propaganda Machine." *Social Media + Society* 2019: 1-13; Freelon, Deen and Tetyana Lokot. 2020. "Russian Twitter disinformation campaigns reach across the American political spectrum." *The Harvard Kennedy School Misinformation Review* 1(1): 1-9; Linvill, Darren L., Brandon C. Boatwright, Will J. Grant, and Patrick L. Warren. 2019. "'THE RUSSIANS ARE HACKING MY BRAIN!' investing Russia's internet research agency twitter tactics during the 2016 United States presidential campaign." *Computers in Human Behavior* 99: 292-300; Linvill, Darren L. and Patrick L. Warren. 2020. "Engaging with Others: How the IRA Coordinated Information Operation Made Friends." *The Harvard Kennedy School Misinformation Review* 1(2): 1-14; Linvill, Darren L. and Patrick L. Warren. 2020. "Troll Factories: Manufacturing Specialized Disinformation on Twitter." *Political Communication*: 1-21; Zhang, Yini, Josephine Lukito, Min-Hsin Su, Jiyoun Suk, Yiping Xia, Sang Jun Kim, Larissa Doroshenko, and Chris Wells. 2021. "Assembling the Networks and Audiences of Disinformation: How Successful Russian IRA Twitter Accounts Built Their Followings, 2015-2017." *Journal of Communication* 71(2): 305-331.

[8] Lukito, Josephine, Jiyoun Suk, Yini Zhang, Larissa Doroshenko, Sang Jung Kim, Min-Hsin Su, Yiping Xia, Deen Freelon, and Chris Wells. 2019. "The Wolves in Sheep's Clothing: How Russia's Internet Research Agency Tweets Appeared in U.S. News as Vox Populi." *The International Journal of Press/Politics* 25(2): 196-216.

users that spread Internet Research Agency content,[9] and the timing of Internet Research Agency posts across different social media platforms.[10]

Policymakers, private companies, and mass publics have all clearly caught on to the importance of influence operations. However, International Security Studies lags behind. Five years after Zuckerberg's wake up call, International Relations scholars have yet to have such a reckoning of their own. Our theories of public opinion formation and the role of public opinion in elite decision-making are primarily from a pre-social media age, and basic questions about the motivations and mechanisms of influence operations remain undertheorized or underexplored. In this work, I will contribute to answering these questions.

How can we conceptualize foreign influence operations? Is pretending to be a member of another state to influence foreign citizens' views about external affairs a viable strategy? Does plausible deniability afford states the benefits theorized by existing covert action literature? International Security Studies offers few answers, which are necessary for a deeper understanding of the psychological mechanisms and potential effects of influence operations in the cyber-age—the central scholarly goal of this dissertation. Such an understanding contributes to the academic study of secrecy and to the broader imperative to conceptualize the threat influence operations pose to democratic societies. In turn, it can then form the basis for devising strategies to effectively counter foreign influence operations.

The Contribution in Brief

In this dissertation, I make three distinct contributions to understanding the psychological mechanisms and effects of foreign influence operations, corresponding to the three papers that form the core of my dissertation: first, I engage in concept-building to better conceptualize what qualifies as an influence operation, why influence operations are prevalent today, and why International Security Studies scholars should care about the phenomenon; second, I show that, contrary to the conventional wisdom in International Relations, peer influences can affect citizens' foreign policy attitudes as much as elite cues and thus influence operations that falsify peer or public cues may be effective; third, I show that governments that outsource influence

[9] Badawy, Adam, Kristina Lerman, and Emilio Ferrara. 2019. "Who Falls for Online Political Manipulation?" In *Companion Proceedings of the 2019 World Wide Web Conference (WWW'19 Companion)*: 162-168.
[10] Lukito, Josephine. 2019. "Coordinating a Multi-Platform Disinformation Campaign: Internet Research Agency Activity on Three U.S. Social Media Platforms, 2015 to 2017." *Political Communication* 37(2): 238-255.

operations to private firms can benefit from plausible deniability and do not forego the supposed benefits of getting caught. In this section, I will briefly introduce the three papers and how the three discrete contributions allow me to answer the central question in detail.

My first dissertation paper, "Foreign Influence Operations and International Security Studies" is a concept-building paper. To better understand the psychological mechanisms and effects of influence operations, we first need a stronger conception of what influence operations are and how they differ from other forms of statecraft. I provide a conceptual exegesis of influence operations; by constructing a definition in component parts, I attempt to draw clear boundaries around what qualifies as an influence operation. I argue that foreign influence operations are prevalent, and likely to remain so, because they are cost effective, difficult to deter, and because propagandists are prone to inflate their impact. As a result, influence operations will remain a challenge for policymakers and will continue to provide opportunity for Security Studies scholars to explain a salient phenomenon.

My first dissertation paper motivates the need to better understand why states opt for covert informational campaigns (instead of overt ones).[11] Scholars have recently made great advances in the study of secrecy and covert action, but this work overwhelmingly focuses on kinetic forms of covert action, not non-kinetic forms of covert action (i.e., influence operations). The covert action literature suggests that secrecy offers the perpetrator two categories of benefits—increasing operational effectiveness and limiting political blowback. My two empirical dissertation papers contribute to the study of these two benefits of secrecy as applied to influence operations.

First, a perpetrator might choose the covert route out of operational necessity to increase the effectiveness of an operation. For a historical example, in 480 BCE Athenians used a covert mission at the Battle of Salamis. As O'Driscoll writes:

> Themistocles used disinformation to induce the Persians to seek battle at a time and place that was to their disadvantage. He sent his trusted slave, Sikinnos, masquerading as a traitor, to inform the Persians that the Greek fleet was riddled with discord and intending to flee as soon as darkness next fell. The Persians took the bait and launched an attack that played into Athenian hands.[12]

[11] The covert criterion of influence operations is based on my definition, but it is also common to a number of other definitions of influence operations. E.g. Nimmo, Ben. 2020. "The Breakout Scale: Measuring the Impact of Influence Operations." *Brookings Institution Report*.

[12] O'Driscoll, Cian. 2015. "Rewriting the Just War Traditions: Just War in Classical Greek Political Thought and Practice." *International Studies Quarterly* 29: 1-10.

State-backed propagandists today often follow what we might call the "Sikinnos Strategy": assuming a false identity to spread information abroad more credibly. Sikinnos pretended to be a traitor; personas used in influence operations often pretend to be members of the target state. In their seminal work on propaganda, Jowett and O'Donnell call this a "Deflective Source Model," wherein the propagandist uses a deflective source, which becomes the apparent source of the message.[13]

The notion that states benefit from hiding their identity when waging influence operations is ostensibly intuitive. Research in psychology suggests that the persuasiveness of a message is dependent on the perceived credibility of a source.[14] Since people exhibit in-group out-group biases, they will be more receptive to information coming from their peers or other in-group members.[15] So, foreign actors may try to hide their identity when conveying information to audiences. However, cue-taking literature in International Relations suggests that citizens form foreign policy preferences by following their favored elites. This top-down conventional wisdom creates a puzzle for influence operations scholars: Why would foreign actors try to influence people's foreign policy preferences via falsified peer and public cues if citizens merely follow the top? Can peer or public cues influence citizens' foreign policy preferences?

To determine whether the mechanism of peer cues is a viable strategy for influence operations that seek to change foreign policy preferences in the target state, in Paper 2 I conduct four survey experiments on nationally representative samples in the United States and United Kingdom where I vary public and expert support for a retaliatory strike to a cyber-attack. If the top-down conventional wisdom holds, we would expect elite cues to significantly outweigh public ones. However, I do not find this to be the case. Both elite and peer cues impact people's policy preferences. This suggests that influence operations that operate via peer and public cues may be effective in shifting foreign policy preferences; in other words, the "operational effectiveness" logic for secrecy in influence operations holds because peer and public cues can

---

[13] Jowett, Garth S. and Victoria O'Donnell. 2012. *Propaganda and Persuasion: Fifth Edition.* SAGE Publications, Inc. p. 25. For a discussion of the deflective source model in application to a contemporary influence operation, see: Golovchenko, Yevgeniy, Cody Buntain, Gregory Eady, Megan A. Brown, and Joshua A. Tucker. 2020. "Cross-Platform State Propaganda: Russian Trolls on Twitter and YouTube during the 2016 U.S. Election." *The International Journal of Press/Politics* 25(3): 357-389.

[14] Baum, Matthew A. and Tim Groeling. 2009. "Shot by the Messenger: Partisan Cues and Public Opinion Regarding National Security and War." *Political Behavior* 31: 157-186; Hovland, Carl I. and Walter Weiss. 1951. "The Influence of Source Credibility on Communication Effectiveness." *Public Opinion Quarterly* 15(4): 635.

[15] See, for example: Maoz, Ifat, Andrew Ward, Michael Katz, and Lee Ross. 2002. "Reactive Devaluation of an 'Israeli' vs. 'Palestinian' Peace Proposal." *Journal of Conflict Resolution* 46(4): 515-546.

influence citizens' foreign policy preferences. Although it is not dispositive, it is a crucial test of a common first-order mechanism of influence.

The second rationale often observed as motivating secrecy for covert action is a desire to limit reputational damage or to avoid escalatory blowback from the target. This rationale is not about operational effectiveness at time T, but about repercussions of an operation at time T+1. In Paper 3, I apply this rationale for secrecy to influence operations and challenge the conventional wisdom. Instead of positing plausible deniability as an unequivocal benefit, I suggest that there may be tradeoffs between the supposed benefits of plausible and implausible deniability. Literature in covert action suggests that states may pursue the covert route to avoid the perception that they are violating international laws or norms and to reduce the likelihood of escalatory pressure in the target state.[16] Yet, research also suggests that states may benefit from *implausible* deniability—when they deny conducting a covert operation even when such a denial is implausible.[17] Cormac and Aldrich write that a state waging implausibly deniable operations would appear omnipotent and stir paranoia in a target state.[18] No studies to date, however, have systematically tested the benefits of waging plausibly deniable versus implausibly deniable influence operations.

How do citizens in a target state react to influence operations that are plausibly versus implausibly deniable? Do states reap the benefits theorized by the kinetic covert action literature? To fill this gap, I present American respondents with a vignette in which there is a social media influence operation during the 2024 U.S. presidential election. I vary whether the operation was conducted by a foreign government, a marketing firm in a foreign country, or an unknown actor, and, if it was conducted by a marketing firm, whether the marketing firm was paid by the foreign government, foreign private citizens, or an unknown entity. States often outsource disinformation work, at least in part, to benefit from plausible deniability.[19] This

[16] Carnegie, Allison. 2021. "Secrecy in International Relations and Foreign Policy." *Annual Review of Political Science* 24: 213-233; Carson, Austin. 2018. *Secret Wars: Covert Conflict in International Politics*. Princeton, New Jersey: Princeton University Press; Poznansky, Michael. 2020. "Revisiting plausible deniability." *Journal of Strategic Studies*: 1-23.
[17] Cormac, Rory and Richard J. Aldrich. 2018. "Grey is the new black: covert action and implausible deniability." *International Affairs* 94(3): 477-494.
[18] Ibid.
[19] Grossman, Shelby and Khadeja Ramali. "Outsourcing Disinformation." *Lawfare*. December 13, 2020. Accessed: https://www.lawfareblog.com/outsourcing-disinformation; DiResta, Renée, Shelby Grossman, and Alexandra Siegel. "In-House vs. Outsourced Trolls: How Digital Mercenaries Shape State Influence Strategies." *Working Paper*.

creates an externally valid study of the extent to which public reactions differ based on whether a foreign government can plausibly deny sponsorship.

I find evidence for the theorized benefits of plausible deniability: when a government's involvement in an influence operation that originated in its country is not known, the state is less likely to face reputational backlash and citizens in the target state are less supportive of retaliation. Surprisingly, I do not find evidence of implausible deniability. Americans do not perceive an operation as more effective or believe the perpetrator country is more influential when an operation is attributed to the government compared to when the government's role is not known. The fact that states benefit from plausible deniability but not implausible deniability suggests that the likely calculation of recent political actors outsourcing their disinformation campaigns to third-party marketing firms is correct. I add the novel insight that assignment of blame and support for retaliation heavily depend on the country of origin: the treatment effects of plausible versus implausible deniability are stronger for allies than adversaries, and respondents are much more likely to blame an adversarial government for operations originating on its soil—even when the government's role is not known.

This dissertation seeks to make a stepwise contribution to understanding the mechanisms and effects of influence operations—a question that has both academic and policy implication. By constructing a definition of influence operations, distinguishing influence operations from related concepts, and showing why they are prevalent, my first dissertation paper lays the foundation for an emerging academic field. My second paper explores a mechanism of influence operations—peer influence—and finds that citizens are as influenced by peer cues as elite ones. This challenges the dominant International Relations literature and draws practical attention to the importance of creating barriers for foreign propagandists who seek to falsify their identity to influence domestic debate. By relying on a cyber-attack vignette, I also contribute insight into how Americans will respond to cyber-attacks with kinetic implications. My third paper makes multiple contributions to a relatively unexplored component of influence operations: how uncertainty over the perpetrator of an operation impacts the target. This study concludes that some potential perpetrators are likely to benefit from plausible deniability while others are not, and highlights the need to raise costs for marketing firms who sell influence-for-hire services.

My dissertation neither claims nor seeks to establish the final word on effects. No one has a simple answer to how much impact these operations have. As Ben Nimmo writes in a recent

report, "One of the greatest challenges in the study of disinformation and influence operations (IOs) is measuring their impact."[20] With all of the research done on Russian interference in 2016, scholars still disagree over whether the operation was unlikely to have "made any difference worth caring about," as Yochai Benkler argues, or whether they "*likely*" had a decisive impact on the outcome, as Kathleen Hall Jamieson concludes.[21] The question of measuring impact is riddled with complexities and confounding variables. Data collection is difficult. Disentangling the effects of foreign influence operations from prior beliefs or other actions is not always possible. Nevertheless, I hope these studies will spur further research on the mechanisms of influence operations and the role of secrecy and plausible deniability in foreign, covert informational campaigns.[22] As I describe in the conclusion of this chapter, democratic resilience depends in part on how citizens and policymakers respond to changes in the information environment and associated threats like cyber-enabled influence operations.

The remainder of this chapter is structured as follows. First, I introduce the reader to the global nature of social media influence operations, relying on a dataset of all operations publicly attributed by Facebook and Twitter. Second, I outline existing literature on public opinion and foreign policy. Two aspects of this literature are particularly relevant: (a) the notion that public opinion about foreign policy is malleable, and (b) the finding that public opinion affects elite decision-making. The fact that public opinion about foreign policy is malleable suggests that influence operations could possibly affect citizens' preferences about external affairs. The fact

---

[20] Nimmo 2020. See also: Andrew and Mitrokhin 2001.

[21] Jamieson, Kathleen Hall. 2020. *Cyberwar: How Russian Hackers and Trolls Helped Elect a President: What We Don't, Can't, and Do Know*. 2020. Oxford University Press. p. 34; Benkler, Yochai. 2020. "The Danger of Overstating the Impact of Information Operations." *Lawfare*. Accessed: https://www.lawfareblog.com/danger-overstating-impact-information-operations. See also: Benkler, Yochai, Robert Faris, and Hal Robert. 2018. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford University Press. See also: Bail, Christopher A., Brian Guay, Emily Maloney, Aidan Combs, D. Sunshine Hillygus, Friedolin Merhout, Deen Freelon, and Alexander Volfosky. 2020. "Assessing the Russian Internet Researcn Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017." *Proceedings of the National Academy of Sciences* 117(1): 243-250; Yaffa, Joshua. "Is Russian Meddling as Dangerous As We Think?" *The New Yorker*. September 7, 2020. Accessed: https://www.newyorker.com/magazine/2020/09/14/is-russian-meddling-as-dangerous-as-we-think; Goldstein, Josh A. and Renée DiResta. "Foreign Influence Operations and the 2020 Election: Framing the Debate." *Lawfare*. October 23, 2020. Accessed: https://www.lawfareblog.com/foreign-influence-operations-and-2020-election-framing-debate

[22] For a review of studies that consider effects, see Bateman, Jon, Elonnai Hickock, Laura Courchesne, Isra Thange, and Jacob N. Shapiro. 2021. "Measuring the Effects of Influence Operations: Key Findings and Gaps from Empirical Research." *Carnegie Endowment for International Peace.* Of note, this paper does not distinguish between overt and covert operations or define what they include under "influence operations." The majority of the 84 studies they review are overt in nature and range from study the effects of exposure to COVID-19 misinformation to various aid programs.

that public opinion affects elite decision-making suggests that a change in public views could affect state behavior. Neither of these literatures makes conclusive statements about the role of influence operations—which they do not address directly. Rather, they create the possibility that influence operations *could* have effects, motivating the need for further study. Further, each of these two dynamics is impacted by the creation of the Internet and the rise of social media, though International Relations scholarship has not fully taken these developments into account. As I describe later in this chapter, public opinion may have a larger effect on elite decision-making in the cyber age for three reasons: the public has access to more information about foreign affairs due to social media, public opinion (or the appearance of public opinion) is more easily known to elites via social media, and because social media can lower the bar to collective action—raising the costs for elites who fail to take public opinion into account. Finally, I conclude the chapter with implications of foreign influence operations for the resilience of democracy before turning to the three dissertation papers.

The Global Nature of Cyber-Enabled Influence Operations

The study of influence operations is particularly important today because of the prevalence and political salience of social media influence operations. Today, social media influence operations are used by a diverse set of actors across the globe. This is evident when examining the networks of accounts publicly removed by Facebook and Twitter from their respective platforms.[23]

When Facebook or Twitter removes a network of accounts for involvement in an influence operation, the platforms make a public announcement of doing so. In June 2021, Facebook published a dataset of all such announcements from 2017-2021. This includes whether the operation was domestic, foreign, or both domestic and foreign (i.e. if it targets populations in

[23] These networks are some of the clearest evidence of contemporary influence operations, in part because the social media companies have access to back-end data such as IP addresses that allows them to make attributions that open-source researchers alone could not. For examples of academic research studying the networks of accounts removed from Facebook or Twitter in what the industry calls "takedowns," see: Bail et al. 2020; Barrie, Christopher and Alexandra Siegel. 2021. "Kingdom of Trolls? Influence Operations in the Saudi Twittersphere." *Journal of Quantitative Description: Digital Media* 1: 1-41. Grossman, Shelby, Katie Jonsson, Nicholas Lyon, and Lydia Sizer. "Slanted Narratives, Social Media, and Foreign Influence in Libya." *Working Paper*; Linvill et al. 2019; Linvill and Warren 2020; Lukito, Josephine. 2020. "Coordinating a Multi-Platform Disinformation Campaign: Internet Research Agency Activity on Three U.S. Social Media Platforms, 2015 to 2017." *Political Communication* 37(2): 238-255; Steward, Leo G., Ahmer Arif, and Kate Starbird. 2018. "Examining trolls and polarization with a retweet network." *MIS2, 2018.*

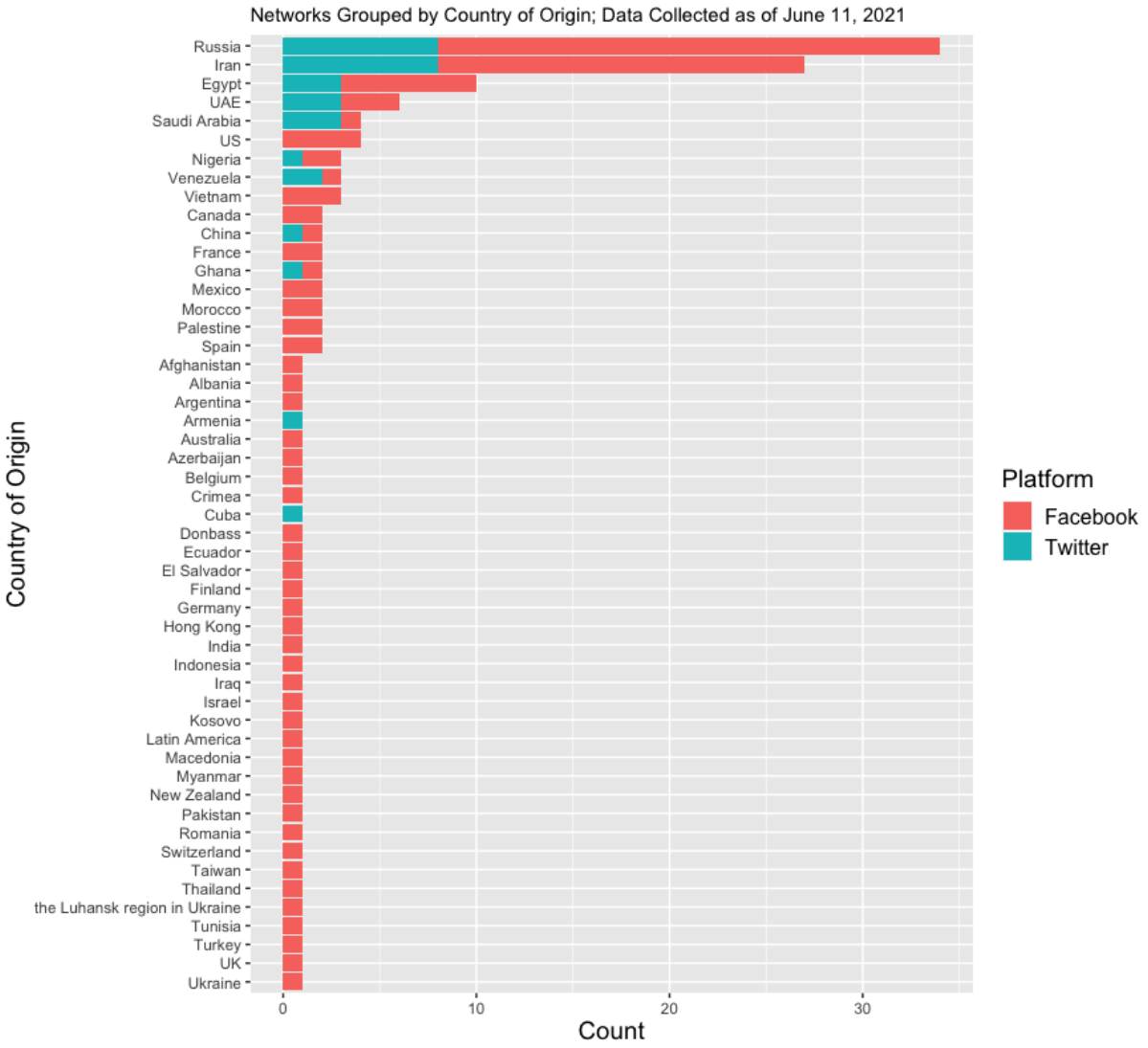
the country of origin as well as external countries).[24] Although Twitter has yet to publish a coded spreadsheet of operations, they make regular announcements and post content from influence operations to a centralized online archive.[25] Thus, I was able to code whether operations on Twitter were domestic or foreign in nature as well.

In Figure 1, I show the number of foreign influence operations that, according to Facebook or Twitter, originated in each country. Foreign influence operations on Facebook and Twitter have originated in 51 different countries. Russia is the most frequent country of origin of foreign influence operations on the two platforms, followed by Iran and Egypt.

[24] Gleicher, Nathaniel, Margarita Franklin, David Agranovich, Ben Nimmo, Olga Belogolova, and Mike Torrey. 2021. "The State of Influence Operations 2017-2020" *Facebook Report.* p. 43. Accessed: https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf

[25] To access this archive, visit: https://transparency.twitter.com/en/reports/information-operations.html

Figure 1: Networks Removed by Facebook or Twitter for Involvement in Foreign Influence Operations



Three caveats to this dataset and Figure 1 are worth noting. First, this data only includes networks that are found by the platforms or independent researchers and are publicly removed. As a result, this is likely a significant undercount of networks of accounts engaged in foreign influence operations on these two platforms. The fact that some operations recently removed began operating years ago suggests that operations evade detection and the total number of operations likely far exceeds the number that have been caught. Nathaniel Gleicher, Head of Security Policy at Facebook, has publicly said, "I think we should be careful about saying that

we know what the denominator is."[26] As a result, our understanding of the scope of the problem is necessarily incomplete.

Second, each data point in Figure 1 reflects a *network* of accounts that have been removed. However, networks are often involved in many operations—targeting people in different countries or used to spread information about different topics or at different time points. For example, in June 2020 Twitter removed 23,750 accounts that it attributed to the People's Republic of China. These accounts focused on diverse topics over time, such as Hong Kong protests, COVID-19, Taiwan, and Geo Wengui, yet they form a single network.[27] Treating the number of networks as a proxy for the number of operations would likely significantly undercount the number of influence operations. A historical analogy may be clarifying: a single agent of influence in the past could be used for many different influence operations; so too can a social media account involved in one operation be used in others. These first two caveats suggest that social media influence operations are far more wide-ranging than even Figure 1 would lead us to believe.

Third, a challenge with coding datasets of takedowns is that it can be difficult to distinguish whether an operation is domestic or foreign, and what it means for an operation to be domestic or foreign.[28] Throughout my dissertation, I refer to foreign operations as operations in which a perpetrator in Country A targets individuals in Country B (or B, C, and D, for example). Sometimes the audience is clear cut, such during the 2020 U.S. Presidential election campaign, when Twitter accounts that were operated from Iran pretended to be Americans, tweeted in English, and tagged U.S.-based news outlets.[29] At other times, however, even if it is clear that actors are engaged in covert, informational campaigns, it can be difficult to determine who they target.

---

[26] Dwoskin, Elizabeth. "Russia is still the biggest player in disinformation, Facebook says." *The Washington Post.* May 26, 2021. Accessed: https://www.washingtonpost.com/technology/2021/05/26/facebook-disinformation-russia-report/

[27] See: Wallis, Jake, Tom Uren, Elise Thomas, Albert Zhang, Samantha Hoffman, Lin Li, Alex Pascoe, and Danielle Cave. 2020. "Retweeting through the great firewall." *Australian Strategic Policy Institute Report No. 33/2020* and Miller, Carly, Vanessa Molter, Isabella Garcia-Camargo, and Renée DiResta. 2020. "Sockpuppets Spin COVID Yarns: An Analysis of PRC-Attributed June 2020 Twitter takedown." *Stanford Internet Observatory Report.*

[28] For a broader discussion of whether and why foreign influence operations are a distinct threat from domestic operations, see: Ördén, Hedvig and James Pamment. 2021. "What Is So Foreign About Foreign Influence Operations?" *Carnegie Endowment for International Peace.* Accessed: https://carnegieendowment.org/2021/01/26/what-is-so-foreign-about-foreign-influence-operations-pub-83706

[29] Goldstein, Josh A. and Lindsay Hundley. 2020. "Viva la #Resistance? Now-Suspended Twitter Accounts from Iran Claim #Resistance Personas in US Politics and Middle Eastern Affairs." *Stanford Internet Observatory Report.*

Public Opinion About Foreign Policy

Before turning to the three core papers of the dissertation in the next three chapters, in the remainder of this chapter, I provide an overview of relevant literatures for understanding the role of public opinion in International Security. The study of influence operations most directly speaks to the question of public opinion formation and its effect on state policy. I pay special attention to how the rise of the Internet affects the relationship between public opinion and elite decision-makers. The conventional wisdom only several decades ago suggested that the public lacked structured beliefs about foreign policy, and that public opinion was not relevant for elite decision-making.[30] If this were the case, then influence operations targeting mass publics would be a waste of resources. However, advancements in Political Science and International Relations since the Vietnam War have challenged these views, suggesting both that public opinion is malleable, and that it matters for elite decision-making and the functioning of a democratic society. In the cyber age, states can not only turn to traditional broadcast modes of influence, but a range of social media and cyber means as well. With newfound abilities to reach foreign publics, the question of the public's malleability and the role of public opinion on elite decision-making and the functioning of a democratic society becomes only more important. Yet, limited work has explored these questions with reference to covert informational campaigns, an increasingly common occurrence in the social media age.

From Almond-Lippmann to the Top-Down Model

In the aftermath of World War II, a scholarly view known as the Almond-Lippmann consensus formed, which was pessimistic about the public's understanding of, and reaction to, foreign policy issues. Lippmann famously wrote that "the unhappy truth" is that "the prevailing public opinion has been destructively wrong at crucial junctures." It has led government to be "too late with too little, too long with too much, too pacifist in peace or bellicose in war," he continued.[31] Almond attributed the mass public's lack of ability to form sound foreign policy views to factors including inattentiveness, sophistication of foreign policy challenges, and state secrecy. In a lecture on "Public Opinion and National Security Policy" Almond said that the

[30] For a discussion of this literature, see: Knopf, Jeffrey W. 1998. "How Rational is "The Rational Public"? Evidence from U.S. Public Opinion on Military Spending." *Journal of Conflict Resolution* 42(5): 544-571.
[31] Lippmann, Walter. 1955. *Essays in the Public Philosophy*. Boston: Little Brown.

public "looks at headlines in the press, or listens to news over the radio for *cues* for *mood* responses in public discussion of foreign policy. It does not listen to the content of discussion but to its tone."[32]

More recent scholarship has categorized the Almond-Lippmann consensus in a few ways. Knopf writes that the consensus contains both an empirical observation and a theory that provides an explanation for that observation. The observation is that the public responds too slowly to foreign threats, and, then, when it does respond, it overreacts. This observation is explained by "mood theory," which posits that the public is typically uninterested and indifferent to foreign policy. When the public realizes there is an external threat, this produces panic, causing exaggerated swings in public views.[33] Holsti summarizes the Almond-Lippmann consensus in three propositions about public opinion:

- "(1) it is volatile and thus provides inadequate foundations for stable and effective foreign policies,
- (2) it lacks coherence or structure, but
- (3) in the final analysis, it has little if any impact on foreign policy"[34]

If the Almond-Lippmann theory accurately described public views about foreign policy, influence operations targeting mass publics about foreign policy issues would be predestined to fail—both because the public's views about foreign policy would not be rational, and because those views would not influence policy.[35] Studies of public opinion about the Vietnam War, however, mounted a significant challenge to the Almond-Lippmann consensus. In a groundbreaking study of public opinion about the Vietnam War, Verba et al. found that public opinion is "relatively orderly. The public expressed concern about the war and was relatively informed about it."[36] Caspary found that the public was relatively stable in its support for internationalism and policies like the Marshall Plan.[37] In his study of the Vietnam and Korean

---

[32] Almond, Gabriel A. 1956. "Public Opinion and National Security Policy." *Public Opinion Quarterly* 20(2): 371-378.
[33] Knopf 1998, pp. 545-546.
[34] Holsti, Ole R. 1992. "Public Opinion and Foreign Policy: Challenges to the Almond-Lippmann Consensus." *International Studies Quarterly* 36: Ibid, 439.
[35] Influence operations targeting mass publics are meaningful subset of influence operations, but not inclusive of all types of influence operations. As I describe further in Paper 1, influence operations can have diffuse targets—like broad publics in other countries—or discrete targets—like ethnic groups, elites, or even individual leaders.
[36] Verba, Sidney, Richard A. Brody, Edwin B. Parker, Norman H. Nie, Nelson W. Polsby, Paul Ekman, and Gordon S. Black. 1967. "Public Opinion and the Vietnam War." *American Political Science Review* 61(2): 317-333.
[37] Caspary, William R. 1970. "The "Mood Theory": A Study of Public Opinion and Foreign Policy." *American Political Science Review* 64(2): 536-547.

Wars, Mueller argues that the American public's support for war is responsive to U.S. soldier casualties.[38] These studies are seminal contributions to a broader literature that shows that public opinion about foreign policy is neither as volatile nor as incoherent as was previously thought.[39]

After moving away from the Almond Lippmann consensus, a puzzle remained: How could Americans hold rational views of foreign policy when they still were relatively uninformed about the details of foreign policy issues? Lack of public knowledge about foreign policy issues is well-documented. For example, during the Ukraine crisis in 2014, few Americans could correctly locate Ukraine on a map.[40] A Gallup-conducted poll in 2019 found that fewer than half of American respondents could identify Afghanistan as the country that provided Al-Qaeda with safe haven before 9/11, despite the fact that the United States has conducted a war in Afghanistan for the decades since as a result.[41] Ignorance may be rational—Downs famously argued that an individual's vote has a negligible impact on the outcome of an election; thus, the cost-benefit analysis regarding whether to take the time to educate one's self on the range of relevant issues favors remaining uninformed.[42] Williams terms this "rational *acquisitional* ignorance."[43] The logic applies to foreign policy perhaps even more so than to voting, since the ramifications of foreign policy decisions are typically less directly felt by Americans than those of economic or domestic policy.[44]

---

[38] Mueller, John E. 1973. *War, Presidents, and Public Opinion*. New York: John Wiley & Sons.

[39] Jentleson, for example, argues that in the post-Cold War era, the American public has been more likely to favor the use of force in response to external threats than as a tool for internal political change in other nations. Jentleson, Bruce W. 1992. "The Pretty Prudent Public: Post Post-Vietnam American Opinion on the Use of Military Force." *International Studies Quarterly* 36(1): 49-73. See also Jentleson, Bruce W. and Rebecca L. Britton. 1998. "Still Pretty Prudent: Post-Cold War American Public Opinion on the Use of Military Force." *Journal of Conflict Resolution* 42(4): 395-417; Nincic, Miroslav. "A Sensible Public: New Perspectives on Popular Opinion and Foreign Policy." *Journal of Conflict Resolution* 36(4): 772-789; Oneal, John R., Brian Lian, and James H. Joyner Jr. 1996. "Are the American People "Pretty Prudent"? Public Responses to U.S. Uses of Force, 1950-1988." *International Studies Quarterly* 40(2): 261-280.

[40] Dropp, Kyle, Joshua D. Kertzer, and Thomas Zeitzoff. 2021. "The less Americans know about Ukraine's location, the more they want U.S. to intervene." *The Monkey Cage.* Accessed: https://www.washingtonpost.com/news/monkey-cage/wp/2014/04/07/the-less-americans-know-about-ukraines-location-the-more-they-want-u-s-to-intervene/

[41] National Geographic, Gallup, and the Council on Foreign Relations. 2019. "U.S. Adults' Knowledge About the World." *Joint Report.* p. 4. Accessed: https://cdn.cfr.org/sites/default/files/report_pdf/NatGeo_CFR_US%20Knoweldge.pdf

[42] Downs, Anthony. 1957. "An Economic Theory of Political Action in a Democracy." *Journal of Political Economy* 65(2): 135-150.

[43] Williams, Daniel. 2020. "Motivated ignorance, rationality, and democratic politics." *Synthese*: 1-21.

[44] Wildavsky writes, "While many people know about the impact of social security and Medicare, few know about politics in Malawi." Wildavsky, Aaron. 1966. "The Two Presidencies." *Society* 35(2): 23-31.

A significant strand of literature post-Almond-Lippmann consensus focused on the role of elite-cues in shaping the public's foreign policy preferences—serving as a solution to the gap between underlying knowledge and rational policy preferences. Citizens can use elites as heuristics to form rational policy preferences even while lacking full knowledge themselves. The dominant model of elite-cue taking suggests that people do not rely on all elites equally, but instead follow their preferred elites.[45] Typically, this is based on partisan identification. Zaller and Berinsky both predict that when elites are united, the public should follow elite cues, and when elites divide, the public should split along partisan lines.[46] Scholars have found top-down effects on American support for military interventions in both hypothetical scenarios and real-world scenarios in Syria and Iran,[47] on support for membership in the EU among Europeans,[48] and on American casualty tolerance through framing effects.[49] Others have shown that the public is more likely to defer to particular types of elites, like those with military experience.[50]

From a normative perspective, scholars have diverged about the desirability of elite-cue taking and the efficiency of the top-down model. On the one hand, elite cues may allow Americans to overcome informational deficits and form reasonable opinions. As Luppia and McCubbins write, "People often *substitute* the advice of others for the information they lack. This substitution can give people the capacity for reasoned choice."[51] On the other hand, if elite cues shape public opinion, the public may be overly susceptible to framing effects, incapable of holding elites accountable for wrongdoing,[52] and elites may mislead the public about foreign policy.[53]

---

[45] Hovland, Carl I. and Walter Weiss. 1951. "The Influence of Source Credibility on Communication Effectiveness." *Public Opinion Quarterly* 15(4): 635-650.
[46] Literature which challenges the findings of the top-down model: Gelpi, Christopher. 2010. "Performing on Cue? The Formation of Public Opinion Toward War." *Journal of Conflict Resolution* 54(1): 88-116.
[47] Golby, James, Peter Feaver, and Kyle Dropp. 2018. "Elite Military Cues and Public Opinion About the Use of Military Force." *Armed Forces & Society* 44(1): 44-71.
[48] Gabel, Matthew and Kenneth Scheve. 2007. "Estimating the Effect of Elite Communications on Public Opinion Using Instrumental Variables." *American Journal of Political Science* 51(4): 1013-1028.
[49] Boettcher III, William A. and Michael D. Cobb. 2006. "Echoes of Vietnam? Casualty Framing and Public Perceptions of Success and Failure in Iraq." *Journal of Conflict Resolution* 50(6): 831-854.
[50] Jost, Tyler and Joshua D. Kertzer. "Armies and Influence: Public Deference to Foreign Policy Elites." *Working Paper.*
[51] Lupia, Arthur and Mathew D. McCubbins. 1998. *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* Cambridge University Press. p. 2.
[52] Agadjanian, Alexander. 2021. "When Do Partisans Stop Following the Leader?" *Political Communication* 38(4): 351-369.
[53] Potter, Philip B. K. and Matthew A. Baum. 2014. "Looking for Audience Costs in all the Wrong Places: Electoral Institutions, Media Access, and Democratic Constraint." *The Journal of Politics* 76(1): 167-181. Mearsheimer, John

The top-down story of public opinion formation in International Relations was always incomplete. Page and Jacob compare the responses of the public and policymakers to identically worded survey questions concerning foreign policy alternatives. Using eight Chicago Council on Foreign Relations surveys between 1974 and 2002 involving a survey of elite "foreign policy leaders" (from the executive branch, House of Representatives, and the U.S. Senate) as well as the public, the scholars find at least a 10 percentage point difference between elite and public preferences on 73% of the 1,153 survey questions asked.[54] The story, however, is particularly incomplete in the cyber age.

Democratization of Influence in the Cyber Age

The rise of the Internet and social media has dramatically shifted the way that people consume information. According to a 2020 Pew poll, 86% of Americans report getting news from a smartphone, computer, or tablet "sometimes" or "often"—more than television and more than double the rate of print publications.[55] About half of Americans report getting news from social media "often" or "sometimes."[56] Research has shown that the Internet has a "competitive displacement effect" on traditional media,[57] and has "endangered the concept of one-way news."[58] Users select from a broader range of sources, peers on social media serve the function of managing editors in deciding what to share—and thereby recommend—to friends, and a range of potential cue givers can easily reach target audiences. As DiResta writes, "Zero-cost publishing eliminated the old editorial gatekeepers and gave anyone a voice."[59] For public

J. 2013. *Why Leaders Lie: The Truth About Lying in International Politics*. Oxford University Press. Schuessler, John M. 2015. *Deceit on the Road to War*. Cornell Studies in Security Affairs.

[54] Page and Jacobs 2006.

[55] Pew Research Center. "Americans more likely to get news on digital devices from new websites, apps and search engines than from social media." *Pew Research Center.* January 11, 2021. Accessed: pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices/ft_2021-01-12_socialmedia_03/

[56] Shearer, Elisa and Amy Mitchell. "News Use Across Social Media Platforms in 2020." *Pew Research Center.* January 12, 2021. Accessed: https://www.journalism.org/2021/01/12/news-use-across-social-media-platforms-in-2020/?utm_source=AdaptiveMailer&utm_medium=email&utm_campaign=21-01-12%20Social%20Media%20Update%20General%20Distro&org=982&lvl=100&ite=7690&lea=1696518&ctr=0&par=1&trk=

[57] Dimmick, John, Yan Chen, and Zhan Li. 2004. "Competition Between the Internet and Traditional News Media: The Gratification-Opportunities Niche Dimension." *The Journal of Media Economics* 17(1): 19-33.

[58] Sagan, Paul and Tom Leighton. 2010. "The Internet & the future of news." *Daedalus* 139(2): 119-125.

[59] DiResta, Renée. 2018. "Computational Propaganda: If You Make It Trend, You Make It True." *The Yale Review* 106(4): 15.

opinion formation about foreign policy issues, foreign political cues, influencers, and peers play an outsized role compared to the past.

The traditional assumption is that unlike domestic political elites, foreign elites should not have a significant effect on American public opinion.[60] The logic of the domestic cue-taking literature hinges on the notion that citizens take cues from intra-party elites because those elites share their interests or a particular social identity. Because foreigners are out-group members, they are unlikely to be seen as credible or trustworthy, and thus have limited ability to persuade foreign publics via direct appeals. Moreover, citizens may view foreign cues as illegitimate or undesirable "foreign interference" in domestic affairs.[61]

Recent studies challenge these assumptions, showing that publics are responsive to foreign political cues. Studies of American public opinion about the Iraq War find that foreign political elites can explain why Americans opposed intervention despite unity among their elites.[62] Other studies focus on whether different American cues have different effects abroad: Schatz and Levine found that a statement attributed to President Bush received less support in Kyrgyzstan and Tajikistan than a statement attributed to an ambassador, ordinary American, or no attribution.[63] Agadjanian and Horiuchi find that attributing a statement to Trump compared to an anonymous Congressman produces some negative effects abroad, but the policy content "has the largest effect in worsening opinion of the U.S."[64]

In the social media era, foreign elites can easily reach mass target audiences; Barberá and Zeitzoff note that by the end of 2014, "more than 76 percent of world leaders had an active Twitter or Facebook account."[65] The authors highlight that social media provides a medium to

---

[60] I thank Tyler Jost for helping to shape my views on foreign political cues and traditional assumptions about their lack of effects in the International Relations literature.

[61] Dragojlovic, Nick. 2015. "Listening to Outsiders: The Impact of Messenger Nationality on Transnational Persuasion in the United States." *International Studies Quarterly* 59: 73, citing Keck and Sikkink 1999, p. 93.

[62] Hayes, Danny and Matt Guardino. 2011. "The Influence of Foreign Voices on U.S. Public Opinion." *American Journal of Political Science* 55(4): 830-850; Murray, Shoon. 2014. "Broadening the Debate about War: The Inclusion of Foreign Critics in Media Coverage and Its Potential Impact on US Public Opinion." *Foreign Policy Analysis* 10(4): 329-350.
Likewise, Dragojlovic finds that the nationality of a cue has an effect on Americans' support for increasing U.S. refugee intake. The identity of the cue-receiver also has an effect—British and French messengers were more persuasive than an American messenger among Democrats, and less persuasive among Republicans (2015).

[63] Schatz, Edward and Renan Levine. 2010. "Framing, Public Diplomacy, and Anti-Americanism in Central Asia." *International Studies Quarterly* 54(3): 855-869.

[64] Agadjanian, Alexander and Yosaku Horiuchi. 2020. "Has Trump Damaged the U.S. Image Abroad? Decomposing the Effects of Policy Messages on Foreign Public Opinion." *Political Behavior* 42(2): 584.

[65] Barberá, Pablo and Thomas Zeitzoff. 2018. "The New Public Address System: Why Do World Leaders Adopt Social Media?" *International Studies Quarterly* 62(1): 121-130. Duncombe suggests that social media and

relay messages to international audiences—"it provides an outlet by which world leaders can communicate directly with citizens all around the world, without the need of traditional news media as an intermediary agent."[66] In a world in which foreign elites can easily reach mass target audiences, foreign cues matter too.

In addition to overt foreign political cues, *covert* foreign cues can reach citizens in other states. As I describe in Paper 1, both state and non-state actors can use influence operations to reach target audiences, capitalizing on the anonymity afforded by the Internet and social media. On the Internet, this may come in the form of fake news sites; since legitimate blogs and news sites are created every day, it's easy to hide in their midst.[67] Fake news sites used in influence operations vary. In a comparative co-authored piece on influence operations across the Middle East, I find at least four types of fake news outlets are recurring:

(1) quasi-real, slanted news outlets, which take direction from an actor hidden from the target;
(2) front media, which often repost content without attribution from other websites;
(3) false franchises, which claim to be regional branches of authentic large news outlets; and,
(4) spoofs of real news outlets, that often engage in typosquatting.[68]

In addition to fake news sites, influence operations use a variety of covert tactics on social media. For one, actors run fake accounts to give the false impression of popular support; a recent investigation found that China's rise on Twitter "has been powered by an army of fake accounts that have retweeted Chinese diplomats and state media tens of thousands of times, covertly amplifying propaganda that can reach hundreds of millions of people…"[69]

At other times actors will covertly organize social media influencers to share their desired message. One campaign hired Nigerian influencers to tweet supportive messages about Alex

---

diplomacy interact on both a top-down plan—from policymakers to foreign publics—and a horizontal one—between state policymakers and their counterparts. See: Duncombe, Constance. 2017. "Twitter and transformative diplomacy: social media and Iran-US relations." *International Affairs* 93(3): 545-562.

[66] Ibid p. 124.

[67] Goldstein, Josh A. and Shelby Grossman. 2021. "How disinformation evolved in 2020." *Brookings TechStream.* Accessed: https://www.brookings.edu/techstream/how-disinformation-evolved-in-2020/

[68] M.A., Renée DiResta, Josh A. Goldstein, and Shelby Grossman. "Middle East Influence Operations: Observations Across Social Media Takedowns." Paper prepared for the Project on Middle East Political Science's 2021 Conference, "Digital Activism and Authoritarian Adaptation in the Middle East."

[69] Kinetz, Erika. "Army of fake fans boosts China's messaging on Twitter." *The Associated Press*. May 28, 2021. Accessed: https://apnews.com/article/asia-pacific-china-europe-middle-east-government-and-politics-62b13895aa6665ae4d887dcc8d196dfc

Saab, who was on house arrest there for money laundering.[70] Another campaign used paid influencers to disparage South African banks.[71] This represents an effort to capitalize on the democratization of potential influential cues, since, in the contemporary information environment, cues beyond top-down political elite ones can influence political preferences.

Finally, peer cues may shape what people believe and the content they consume in the social media era. According to Messing and Westwood, social media affects the media landscape in at least two ways.[72] First, people are presented with a variety of sources in one place, leading consumers to select particular stories instead of individual sources. Second, people utilize endorsements from others because social recommendations impact content people are shown. Studies across disciplines find that peers influence people's perceptions of different media outlets[73] and their selection of news stories to consume.[74] Because people are more likely to believe content from sources they see as trustworthy or credible,[75] politically-motivated actors have incentives to wage influence operations in which they pose as members of a target's community in hopes of persuading them. Freelon and Lokot, for example, highlight how Internet Research Agency accounts operating in the United States pretended to be liberals, conservatives, and Black Americans to provoke anger against oppositional groups.[76] IRA employees likely believe that a liberal American Twitter user would be more persuaded by information coming from an account that looks like an American Twitter user than an account showing the real identity of an Internet Research Agency employee. Of note, only one paper to date has tested the

[70] Silverman, Craig. "Why Do A Bunch Of Nigerian Twitter Influencers Want This Alleged Money Launderer To Go Free? They're Being Paid." *BuzzFeed News.* April 9, 2021. Accessed: https://www.buzzfeednews.com/article/craigsilverman/nigerian-twitter-campaign-alex-saab

[71] Le Roux, Jean. 2021. "Paid influencers and sockpuppets drove social media campaign targeting South African banks." *Medium: DFRLab.* Accessed: https://medium.com/dfrlab/paid-influencers-and-sockpuppets-drove-social-media-campaign-targeting-south-african-banks-28bd04cd343

[72] Messing, Solomon and Sean J. Westwood. 2014. "Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online." *Communication Research* 41(8): 1042-1063.

[73] Turcotte, Jason, Chance York, Jacob Irving, Rosanne M. Scholl, and Raymond J. Pingree. 2015. "News Recommendations from Social Media Opinion Leaders: Effects on Media Trust and Information Seeking." *Journal of Computer-Mediated Communication* 20: 520-535.

[74] Anspach, Nicolas M. 2017. "The New Personal Influence: How Our Facebook Friends Influence the News We Read." *Political Communication* 34: 590-606.

[75] For a review of the literature, see: Pornpitakpan, Chanthika. 2004. "The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence." *Journal of Applied Social Psychology* 34(2): 243-281.

[76] Freelon, Deen and Tetyana Lokot. 2020. "Russian Twitter disinformation campaigns reach across the American political spectrum." *The Harvard Kennedy School Misinformation Review* 1(1): 1-9.

relative effects of elite and peer cues on the public's foreign policy preferences, as I describe in Paper 2.[77]

Although the democratization of the news has made way for influence operations to become an increasingly common tool of statecraft, our theories of public opinion formation about foreign policy have yet to adapt to this new reality. Influence operations present a set of unanswered questions about how people form their opinions in the cyber age, the extent to which different types of influence operations can affect public opinion, and the extent to which knowledge of influence operations can undermine faith in democratic institutions, the news, and even truth writ large. Influence operations have been used during both peace time and during times of crises, like in the 2017 Gulf Crisis blockading Qatar or Turkey's 2019 Operation Peace Spring offensive into Syria.[78] Further research is necessary to situate this empirical phenomenon within the International Security Studies literature and to unpack the potential effects of influence operations.

The Effect of Public Opinion on Foreign Policy

The question of public opinion formation and the role of influence operations would have limited significance for scholars of International Relations if public opinion did not shape foreign policy decisions. This was the assumption of the Almond-Lippmann consensus, as well as a related but distinct literature on the "two presidencies."[79]

However, scholarship has also mounted a significant challenge to the assumption that public opinion about foreign policy is unimportant. In fact, the public's role in shaping or constraining elite and presidential decision-making is both empirically supported and central to several foundational theories of International Security. These developments underscore the need

[77] Kertzer, Joshua D. and Thomas Zeitzoff. 2017. "A Bottom-Up Theory of Public Opinion about Foreign Policy." *American Journal of Political Science* 61(3): 543-558.
[78] Jones, Marc Owen. 2019. "Propaganda, Fake News, and Fake Trends: The Weaponization of Twitter Bots in the Gulf Crisis." *International Journal of Communication* 13: 1389-1415; Grossman, Shelby, Fazil Alp Akis, Ayça Alemdaroğlu, Josh A. Goldstein, and Katie Jonsson. 2020. "Political Retweet Rings and Compromised Accounts: A Twitter Influence Operation Linked to the Youth Wing of Turkey's Ruling Party." *Stanford Internet Observatory Report*.
[79] As Wildavsky writes in "The Two Presidencies," the classic article presenting the thesis: "The United States has one president, but it has two presidencies; one presidency is for domestic affairs, and the other is concerned with defense and foreign policy. Since World War II, Presidents have had much greater success in controlling the nation's defense and foreign policies than in dominating its domestic policies." See: Wildavsky 1966, p. 23. For additional work supporting the thesis, see Milner, Helen V. and Dustin Tingley. 2015. *Sailing the Water's Edge.* Princeton University Press. p. 11 footnote 31.

to better understand influence operations, because public opinion is not only malleable, but it also plays a role in a state's foreign policy decisions.

Evidence that Public Opinion Affects Elite Decision-Making

Groundbreaking experiments on elite decision-makers provide empirical evidence that public opinion about foreign policy affects elite decisions. Tomz, Weeks, and Yarhi-Milo argue that public opinion could affect foreign policy in two ways: either through responsiveness or selection. Responsiveness is the idea that incumbent leaders respond to political preferences of constituents to avoid political costs; selection suggests that the public chooses leaders that align with their views, so representatives selected for office align with public wishes. To test responsiveness, they surveyed 87 current and former members of the Israeli Knesset in 2015. They provided a vignette involving a terrorist attack in Israel from Lebanon, and manipulated whether the Israeli public supported or opposed military action. The authors find a 16% difference in support for military action based on whether the public supports or opposes the action. Building on this research, Chu and Recchia surveyed 101 current Members of the UK House of Commons. The authors presented a subset of the MPs with polling information indicating the UK public's support for continuing military activity in the South China Sea. Under the treatment conditions, MPs shifted in favor of such activities as well.[80]

The advantage of experimental evidence is that it shows a direct effect of public opinion in a controlled environment.[81] However, some could argue that, in practice, elites are unlikely to be affected by public opinion because the public knows little about foreign policy issues or because elites are unaware of public opinion.[82] This rebuttal, however, loses force when we consider that observational data also finds that public opinion affects foreign policy decisions. For one, observational data suggests that changes in public opinion typically precede changes in government policy. Examining public opinion and policy changes in the United States from 1935 to 1979, Page and Shapiro find substantial evidence that there is a high level of congruence

[80] Chu, Jonathan and Stefan Recchia. 2020. "Does Public Opinion affect the Preferences of Foreign Policy Leaders? Experimental Evidence from the UK Parliament." *Working Paper*.
[81] On using elite experiments in IR, see: Dietrich, Simone, Heidi Hardt, and Haley J. Swedlund. 2021. "How to make elite experiments work in International Relations." *European Journal of International Relations* 27(2): 596-621.
[82] Busby and Monten argue that American National Election Studies data shows that "foreign policy concerns were not salient the public through the 1990s." p. 106. Busby, Joshua W. and Jonathan Monten. 2012. "Republican Elites and Foreign Policy Attitudes." *Political Science Quarterly* 127(1): 105-142.

between public opinion and policy, and that opinion typically changes *before* policy changes are made.[83] Pomirchy and Schonfeld create issue-specific constituency opinion estimates broken down by party and Congressional district. The authors, focusing on public opinion on issues of trade, security, and immigration policy, find that the higher levels of overall support for a bill in a constituency, the more likely the legislator is to vote for the bill.[84] Likewise, social media analysis of members of the 113th Congress shows that Congressional leaders typically follow, not precede, online discussion by their constituents.[85] If public opinion is malleable to diverse cues and public opinion can affect state policy, further research in International Relations should investigate means of influence in the social media age.

How Social Media Impacts the Bottom-Up Account

The experimental and observational data establishing the public's effects on elite decision-making bolsters the need to better understand the effects of influence operations. The literature suggests that if influence operations affect public opinion, influence operations could, in turn, affect a state's policy decisions. The introduction of social media, as described above, drastically changes the relationship between citizens, elites, and the media, but scholars have "failed to fully grapple with the implications of these changes."[86] While further research beyond the scope of this dissertation is necessary to investigate how changes in the media landscape affect elite responsiveness to public opinion, I posit three reasons why public opinion may play a particularly important role in the cyber age: (1) social media may shift informational asymmetries about foreign policy away from elite dominance; (2) public opinion is more easily

[83] Page, Benjamin I. and Robert Y. Shapiro. 1983. "Effects of Public Opinion on Policy." *The American Political Science Review* 77(1): 175-190. The authors find that the difference between domestic and foreign policy is not significant, providing evidence that changes in foreign policy preferences among the public impact governmental decisions (p. 182).
[84] Pomirchy, Michael and Bryan Schonfeld. "Does Responsiveness Stop at the Water's Edge? Examining the Foreign Policy Electoral Connection" pp. 16-18. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3436890
[85] Barberá, Pablo, Andreu Casas, Jonathan Nagler, Patrick J. Egan, Richard Bonneau, John T. Jost, and Joshua A. Tucker. 2019. "Who leads? Who follows? Measuring issue attention and agenda setting by legislators and the mass public using social media data." *American Political Science Review* 113(4): 883-901.
[86] Baum, Matthew A. and Philip B. K. Potter. 2019. "Media, Public Opinion, and Foreign Policy in the Age of Social Media." *The Journal of Politics* 81(2): 747.

known on a wider range of issues because of social media; and, (3) publics can more easily mobilize due to social media, raising the costs for elites who ignore the public.[87]

First, social media grants citizens access to firsthand information on a wider range of foreign policy issues.[88] As Chesney and Citron write, "thanks to smartphones, which make it easy to capture audio and video content, and social media platforms, which allow that content to be shared and consumed, people today can rely on their own eyes and ears to an unprecedented degree."[89] Aggregate trends also cross borders: Trending hashtags, for example, can alert citizens in one part of the world to policy challenges across the globe. Emerging information technologies may also limit a government's ability to keep secrets from both domestic and foreign publics—weakening state control over both information collection and the disclosure decision.[90] Public opinion, or opinion among at least a part of the population, is likely to form on a wider range of foreign policy issues in an increasingly interconnected world. Thus, while public opinion may have only been relevant on a small subset of salient issues in the past, social media may widen the scope of foreign issues that garner public attention.

Second, social media provides crucial information about how popular (or not) various policies and positions are. If the logic of responsiveness rests on elite knowledge of public views, then we might expect responsiveness to be stronger on, or applicable to, a wider range of topics if public opinion is more frequently known or accessible. As Zeitzoff writes, "social media generates new data and information."[91] In one study of Israel and Hamas' behavior during the 2012 Gaza Conflict, Zeitzoff created an hourly measure of public support for the conflict using social media data. Using a Bayesian structural vector autoregression, he found that Hamas's and Israel's communication on social media responded to changes in support, and that increases in support for rival actors constrained each actor's conflict intensity.[92] In lower salience issues, too,

[87] Note that Baum and Potter come to a different conclusion, arguing that "the current information and political environments erode constraint by inclining constituents to reflexively and durably back "their" leaders and disapprove of opposition." Ibid p. 747.
[88] This is particularly true in states that don't tightly control the use of social media.
[89] Chesney, Robert and Danielle Citron. 2019. "Deepfakes and the New Disinformation War." *Foreign Affairs*. As cited in: Kreps, Sarah. 2020. *Social Media and International Relations*. Cambridge Elements: Cambridge University Press.
[90] Lin-Greenberg, Erik and Theo Milonopoulos. 2021. "Private Eyes in the Sky: Emerging Technology and the Political Consequences of Eroding Government Secrecy." *Journal of Conflict Resolution* 65(6): 1067-1097.
[91] Zeitzoff 2017, 1982.
[92] Zeitzoff, Thomas. 2018. "Does Social Media Influence Conflict? Evidence from the 2012 Gaza Conflict." *Journal of Conflict Resolution* 62(1): 29-63.

opinion is likely to be accessible. This is evident from the fact that journalists frequently cite social media trends and individual tweets as proxies for public opinion.[93]

Third, messaging services including, but not limited to, social media can hasten collective action, possibly lowering the barrier to entry for anti-regime protest.[94] Early scholarly accounts deemed social media a 'liberation technology' for its ability to "empower individuals, facilitate independent communication and mobilization, and strengthen an emergent civil society."[95] More recent scholarship, accounting for the ways that governments have used the Internet for censorship, opinion manipulation, and intimidation, have been less optimistic than the 'liberation technology' thesis. Nevertheless, social media remains a tool of mobilization: Clarke and Kocak show that Facebook and Twitter facilitated a large, nationwide, and seemingly leaderless protest in Egypt on January 25, 2011, thereby mobilizing the Egyptian uprising's "first movers."[96] Even for leaders who do not fear revolution, social media facilitating political mobilization may lead them to be more attuned to public opinion from the onset.

Public Opinion and the Resilience of Democracy

As Hyde writes, "If the end of the 20th century was defined by the relatively widespread acceptance of democracy, the second decade of the 21st century is marked by concerns about backsliding in new and established democracies alike and by a notable decline in foreign support for democracy around the world."[97] Democracies rest on the ability of citizens to form meaningful opinions and to hold leaders accountable for their actions or inactions.[98] Understanding influence operations and their role on public opinion is thus critical to addressing democratic resilience for the decade to come.

As I describe more fully in Paper 1, influence operations can have both first-order and second-order effects. For the perpetrator, first-order effects are about the specific information of

[93] Molyneux, Logan and Shannon McGregor. "When journalists put tweets in news stories, do they transfer too much power to Twitter?" *Nieman Lab*, March 26, 2021. Accessed: https://www.niemanlab.org/2021/03/when-journalists-put-tweets-in-news-stories-do-they-transfer-too-much-power-to-twitter/
[94] Lynch, Marc. "Media, Old and New." *The Arab Uprisings Explained*. Edited by Marc Lynch, Columbia University Press. pp. 93-109.
[95] Diamond, Larry. 2010. "Liberation Technology." *Journal of Democracy* 21(3): 69-83.
[96] Clarke, Killian and Korhan Kocak. 2020. "Launching Revolution: Social Media and the Egyptian Uprising's First Movers." *British Journal of Political Science* 50: 1025-1045.
[97] Hyde, Susan D. 2020. "Democracy's backsliding in the international environment." *Science* 369: 1192-1196.
[98] Warren, Mark E. 2014. "Accountability and Democracy." *The Oxford Handbook of Public Accountability*. pp. 39-54.

the campaign, while second-order effects are about broader outcomes beyond that informational campaign. Influence operations could change people's minds about the information that is the subject of the campaign. But even if influence operations fail to affect people's views about underlying content of the operation, they could have second-order effects—including undermining democracy's global appeal, corroding belief in truth itself, and polarizing democratic societies.

Illiberal states use influence operations to undermine democracy's global appeal. Sometimes this takes the form of tarnishing the brand of specific democracies, like influence operations emphasizing shortcomings in the United States. At other times, governments wage influence operations against democratic elections, undermining faith in key democratic institutions. Goldsmith and Horiuchi ran survey experiments in Japan and found that information about Russian interference in the United States reduces Japanese citizens' faith in the United States as an ally.[99]

Influence operations can be used to undermine one's views in the notion of truth altogether. In a *New York Times* op-ed, Sykes interviews Garry Kasparov, a Russian chess grandmaster and political dissident. Kasparov argues that "The point of modern propaganda isn't only to misinform or push an agenda. It is to exhaust your critical thinking, to annihilate truth." According to Sykes, "Mr. Kasparov grasps that the real threat is not merely that a large number of Americans have become accustomed to rejecting factual information, or even that they have become habituated to believing hoaxes. The real danger is that, inundated with "alternative facts," many voters will simply shrug, asking, "What is truth?" — and not wait for an answer."[100] Herb Lin has outlined how the corruption of the information ecosystem—through influence operations and other means—compounds existential challenges like nuclear war and climate change.[101]

Influence operations can also be used to subvert democratic societies by heightening political polarization. In a groundbreaking study on the effect of foreign electoral intervention, Tomz and Weeks investigate how Americans respond to different types of foreign intervention in

[99] Goldsmith, Benjamin E. and Horiuchi, Yusaku. "Does Russian Election Interference Damage Support for U.S. Alliances? The Case of Japan" *Working Paper*.
[100] Sykes, Charles J. "Why Nobody Cares the President is Lying." *The New York Times* February 4, 2017. Accessed: https://www.nytimes.com/2017/02/04/opinion/sunday/why-nobody-cares-the-president-is-lying.html
[101] Lin, Herbert. 2019. "The Existential Threat from Cyber-Enabled Information Warfare." *Bulletin of Atomic Scientists* 75(4): 187-196.

elections.[102] The authors find a strong partisan double standard: "Both Democrats and Republicans were more likely to condemn foreign involvement, lose faith in democracy, and call for retaliation when a foreign power sided with the opposition, than when a foreign power aided their own party."[103] In one follow-up project, Tomz and Weeks found a partisan asymmetry in Americans' willingness to punish candidates who seek foreign support in their campaign: while Republican candidates largely get a "pass" from their base, Democratic candidates face significant punishment from both parties.[104] Although Tomz and Weeks are focused on a range of foreign electoral interventions, many of which lack the 'covert' criterion to qualify as influence operations, their work provides strong evidence that citizens faced with information of foreign involvement in their country exhibit strong partisan gaps. This helps to explain polling after the 2016 U.S. election, which found that Democrats viewed Russian interference in the election as much more serious than Republicans did.[105] Further research is necessary to understand whether and how democracy can cope with technological advancements and the accompanying abuses of those advancements, in the cyber age.

[102] Some of these interventions—like a public statement supporting one candidate—lack the covert criterion that would make something an influence operation as defined in this dissertation.
[103] Tomz, Michael and Jessica L. P. Weeks. 2020. "Public Opinion and Foreign Electoral Intervention." *American Political Science Review* 114(3): 857.
[104] Tomz, Michael and Jessica L. P. Weeks. n.d.-a "The Domestic Political Costs of Soliciting Foreign Electoral Intervention." *Working Paper*.
[105] Suffolk polls. See: https://www.suffolk.edu/-/media/suffolk/documents/academics/research-at-suffolk/suprc/polls/national/marginal-and-table/final_june_national_crosstabs_6-28_pdftxt.pdf?la=en&hash=F8CCE8E75ECA9AC39155ADCF0AC9B4CCE9250823 p. 21.

Chapter 2: Foreign Influence Operations and International Security Studies

Dissertation Paper 1

Josh A. Goldstein[1]

Abstract

In the aftermath of Russian interference in the 2016 U.S. election, influence operations have become a central focus among policymakers, foreign policy pundits, and publics alike. Yet, influence operations remain under-explored and under-theorized in the International Security Studies literature. Existing studies diverge over how to define influence operations, and few studies in top journals of Political Science, International Relations, or Security Studies focus on the topic. I engage in concept building by constructing a definition in component parts and showing how influence operations differ from related forms of statecraft. I then argue that Security Studies scholars should pay greater attention to influence operations for two reasons. First, influence operations are policy-relevant and will likely remain so for three reasons: they are cost effective, foreign propagandists are prone to inflate their impact, and they are difficult to deter. Second, studying influence operations can contribute to the academic discipline of Security Studies by helping to expand restrictive notions of state power and challenge dominant models of foreign policy decision-making by problematizing the domestic/foreign divide. I conclude by charting possible avenues for a Security Studies research agenda on influence operations.

[1] For feedback on this project at various stages, I thank Sam Bradshaw, Janina Dill, Renée DiResta, Rob Johnson, Steve Rosen, Andrea Ruggeri, Duncan Snidal, the Stanford Internet Observatory team, and participants of Stanford CISAC's Social Science Research Group. Thanks also to attendees at Oxford's Changing Character of War Centre (CCW) Seminar Series and the Oxford Information Warfare Working Group.

Introduction

During the height of the Cold War, the Soviet Union Committee for State Security (the KGB) and East German Foreign Intelligence deployed a series of influence operations to shape the anti-nuclear and global peace movements. Their objective was political—to prevent NATO from deploying "qualitatively new atomic medium-range ballistic missiles by the year 1983"[2]—and their use of covert, informational campaigns, or influence operations, was wide-ranging.[3] In one line of effort, East German Foreign Intelligence covertly supported and directed an unwitting group of former NATO generals campaigning for disarmament. The leader of the so-called "Generals for Peace" was a Stasi agent of influence.[4] The "Generals for Peace" attracted widespread attention in Western press, though their connection to Eastern security agencies was hidden from the public eye.[5] In another line of effort, two leaders of Soviet front groups were active participants in planning a peace demonstration for June 12, 1982, at the same time as the UN Second Special Session on Disarmament was underway. The agents of influence were successful in persuading the demonstration "to direct the focus of the demonstration exclusively against U.S. nuclear weapons systems and away from Soviet weapons systems."[6] Moscow's front organizations pushed the slogan "No New Missiles in Europe"—accepting, by implication, recently deployed Soviet SS-20 medium-range missiles in Europe while opposing U.S. modernization.[7] In *The Mitrokhin Archive*, a key source on Soviet active measures, Andrew and Mitrokhin write that it is reasonable to assume that KGB active measures encouraged the

---

[2] Rid, Thomas. 2020. *Active Measures: The Secret History of Disinformation and Political Warfare*. Farrar, Straus and Giroux. p. 269.

[3] In this paper, as I describe below, I define influence operations as covert informational campaigns intended to affect attitudes or opinions in a manner favorable to the perpetrator. I focus on a particular subset of influence operations: influence operations for political ends that target people in other countries.

[4] An agent of influence is defined as "a person who uses his or her position, influence, power, and credibility to promote the objectives of a foreign power in ways unattributable to that power. Influence operations may be carried out by controlled agents (persons who are recruited, and advance the interests of a foreign power in response to specific orders); "trusted contacts" (persons who consciously collaborate to advance the objectives of a foreign power, but who are not formally recruited and controlled); and unwitting but manipulated individuals. (Schultz, Richard H. and Roy Goodson, ed. *Desinformatsia: Active Measures in Soviet Strategy* New York, N.Y.: Pergamon-Brassey 1984, p. 193-194. as cited by George, Robert K. 1987. "An Historical Investigation of Soviet Strategic Deception" Report no. 87-0955 *Maxwell AFB: Air Command and Staff College*.

[5] For an example of coverage of the "Generals for Peace", see NYT reporting in 1982 on the Generals for Peace at the U.N.: https://www.nytimes.com/1982/06/27/world/generals-for-peace-push-case-at-un.html

[6] United States Department of State. 1986. Active Measures: A Report on the Substance and Process of Anti-U.S. Disinformation and Propaganda Campaigns, 26. Accessed: http://insidethecoldwar.org/sites/default/files/documents/Soviet%20Active%20Measures%20Substance%20and%20Process%20of%20Anti-US%20Disinformation%20August%201986.pdf

[7] Rid 2020.

"overconcentration by most Western peace activists on the nuclear menace posed by Reagan and his NATO allies rather than on that from the Soviet Union."[8]

In the last several decades, files from defectors from the Soviet Union and declassified records in liberal democracies have shown the extent to which the Soviet Union waged influence operations for political ends in other countries;[9] as Thomas Rid testified before the U.S. Senate Select Committee on Intelligence, "The Cold War likely saw more than 10,000 active measures across the world."[10] These operations consumed significant state resources: by early 1985, Soviet active measures had an annual budget between $3 billion and $4 billion—"an estimate that CIA analysts called 'conservative.'"[11] Some of these operations—like the covert Soviet campaign to spread the falsehood that the United States created AIDS—are well known to the public.[12] As late as 2005, 27% of African Americans still believed the Soviet-created conspiracy that AIDS was produced in a government lab.[13] Other influence operations, like Soviet forgeries of U.S. government documents to persuade leaders in developing countries that the CIA and other American agencies had hostile plans against them, are less well known.[14]

While the Soviet Union used influence operations extensively, state-backed influence operations are hardly a tool used exclusively by the Soviet Union or Russia. In World War II, German military intelligence, the Abwehr, spread rumors in Russia that Hitler was not preparing to invade, but rather to issue an ultimatum to Stalin. According to Andrew and Mitrohkin, "It was this illusory threat of an ultimatum, rather than the real threat of German invasion, which increasingly worried Stalin during the few weeks and days before BARBAROSSA."[15] In the same broader conflict, the British Security Coordination planted stories in American news outlets, ran leading polling questions under front organizations to "show" Americans that they

---

[8] Andrew, Christopher and Vasili Mitrokhin. 2001. *The Sword and the Shield: The Mitrokhin Archive and the Secret History of the KGB.* New York: Basic Books. p. 484.
[9] E.g. Andrew and Mitrokhin. *The Sword and the Shield.*
[10] *Disinformation: A Primer in Russian Active Measures and Influence Campaigns Panel II*, Hearing before the Select Committee on Intelligence of the United States Senate, 105th Cong., 1st sess., March 30, 2017, 19. Accessed: https://www.justsecurity.org/wp-content/uploads/2019/07/Disinformation-A-Primer-in-Russian-Active-Measures-and-Influence-Campaigns-Part-2-Panel-II.pdf
[11] Rid 2020, 313.
[12] A major history of the KGB includes that in the "first six months of 1987 alone the story received major coverage in over forty Third World countries." Andrew, Christopher and Oleg Gordievsky. 1990. *KGB The Inside Story.* New York: HarperCollins Publishers. p. 630.
[13] As cited by DiResta, Renee, Michael McFaul, and Alex Stamos. "Here's how Russia will attack the 2020 election. We're still not ready." *The Washington Post*. November 15, 2019.
[14] United States Department of State, "Active Measures."
[15] Andrew and Mitrohkin, *The Sword and the Shield,* 93.

favor intervention, and launched a covert smear campaign against isolationist Congressman Henry Fish.[16] During the Vietnam War, the United States used black and grey radio stations, rumor campaigns, and forgeries to target the North Vietnamese will to fight.[17] Influence operations are a tool of statecraft that have been used by autocracies and democracies, by major powers and small ones, in world wars and in times of relative peace.

In the present era, cyber-enabled influence operations have emerged as a major policy challenge.[18] The Russian Internet Research Agency's (IRA) covert social media efforts to sow "discord in the U.S. political system" leading up to the 2016 U.S. election is the most well-known case of a social media influence operation.[19] But in the last few years alone, hundreds of networks of accounts on Facebook and Twitter have been removed for their use in influence operations. As of June 2020, the Disinfodex project tracked 256 distinct networks that were removed and publicly announced by Facebook, Twitter, or YouTube.[20] This is likely a significant undercount of the number of social media influence operations, given that the same assets are often used for multiple influence operations. Furthermore, the number of takedowns only includes what the social media companies have discovered and chosen to reveal to the public. Martin, Shapiro, and Ilhardt count 76 foreign social media influence efforts targeting 30 different countries from 2013-2019, only including in their dataset operations that produce content designed to appear indigenous to the target state.[21] Social media influence operations may be a relatively new political phenomenon—Facebook, YouTube, and Twitter were only created in 2004, 2005, and 2006, respectively—but they have quickly become a wide-ranging one.

Like influence operations in the past, contemporary state-backed foreign influence operations have both first- and second-order effects. In the first order, these operations can

---

[16] Usdin, Steve. "When a Foreign Government Interfered in a U.S. Election—to Reelect FDR." *Politico Magazine*. January 16, 2017. Accessed: https://www.politico.com/magazine/story/2017/01/when-a-foreign-government-interfered-in-a-us-electionto-reelect-fdr-214634

[17] See Kissinger's memo: https://www.cia.gov/library/readingroom/docs/LOC-HAK-512-7-21-2.pdf

[18] As I describe later in the paper, not all influence operations have political objectives. I am focused on a subset of influence operations—foreign political influence operations—and provide a full exposition in Section I.

[19] As described by Volume I of the Report On the Investigation Into Russian Interference In The 2016 Presidential Election, Special Counsel Robert S. Mueller III. Accessed: https://www.justice.gov/archives/sco/file/1373816/download

[20] For an updated list, see: https://disinfodex.org/

[21] This only includes social media operations that produce content designed to appear indigenous to the target state, leading to a significant undercount of all influence operations. Martin, Shapiro, and Ilhardt 2020.

change people's views or the information environment by, for example, spreading rumors about a government official or making a third-party state appear threatening.[22] Even if the thesis of an influence operation does not convince a target, influence operations can still have second-order effects. They can stir paranoia that influence operations are ever-present, corrode faith in news and elections, and erode common knowledge—knowledge that is agreed upon and forms the foundation of civil society and debate.[23] As cybersecurity expert Herb Lin has written, "the Enlightenment established reason and reality as the foundational pillars of civilized discourse…Cyber-enabled information warfare provides the tactics, tools, and procedures—in short, the means—to replace the pillars of logic, truth, and reality with fantasy, rage, and fear" (2019, 194).[24] Compared to influence operations in the past, however, social media influence operations are cheaper, anonymity is more easily achieved, and information can spread with greater velocity.

The policy community has begun to recognize the renewed threat of foreign influence operations in the digital age. In recent years, the U.S. State Department created a Global Engagement Center to study the threat of foreign influence, the Office of the Director of National Intelligence is in the process of creating a Foreign Malign Influence Center to serve as a hub for information on election interference, and bipartisan Senators have introduced legislation to combat foreign political influence operations.[25] The Carnegie Endowment for International Affairs' Partnership for Countering Influence Operations (PCIO) reviewed 84 policy papers for

---

[22] For example, Twitter accounts from Russia heavily pushed the hashtag, #FireMcMaster. National Security Adviser H.R. McMaster was a Trump administration official who supported the US remaining engaged in NATO and with the EU, contrary to Russia's interests. See: https://www.npr.org/2017/08/20/544817844/russian-bots-tweeting-calls-to-fire-mcmaster-former-fbi-agent-says

[23] Farrell, Henry and Bruce Schneier. 2018. "Common-Knowledge Attacks on Democracy." *Berkman Klein Center Research Publication No. 2018-7*. Tomz, Michael and Jessica L. P. Weeks. 2020. "Public Opinion and Foreign Electoral Interference." *American Political Science Review* 114(3): 857.
For an in-depth study of what constitutes successful covert action, see: Cormac, Rory, Calder Walton, and Damien Van Puyvelde. 2021. "What constitutes successful covert action? Evaluating unacknowledged interventionism in foreign affairs." *Review of International Studies* 1-18.

[24] Lin, Herbert. 2019. "The existential threat from cyber-enabled information warfare." *Bulletin of the Atomic Scientists* 75(4): 187-196.
Emily O. Goldman, cyber strategist at U.S. Cyber Command and the National Security Agency, writes: "Most prominently, state actors are conducting disinformation campaigns and information manipulation in order to weaken domestic political cohesion and confidence in government institutions. These threats demand an immediate response." Goldman, Emily O. 2020. "From Reaction to Action: Adopting a Competitive Posture in Cyber Diplomacy." *Texas National Security Review* 3(4): 84-101.

[25] For example, see the "Countering the Chinese Government and Communist Party's Political Influence Operations Act" co-sponsored by Democratic and Republican Senators. https://www.congress.gov/bill/116th-congress/senate-bill/480/all-info

best practices in combatting influence operations, showing the widespread effort by think tankers to chart a path forward.[26] However, despite the contemporary prominence and historical importance of influence operations, the subject is relatively unexplored and untheorized in International Relations and Security Studies. Few studies in top journals examine influence operations; studies in related disciplines and the think tank community have implicit disagreements over how to define the subject and what should (and what should not) be included in the category of interest.[27] The failure of Political Science, International Relations, and even Security Studies to interrogate influence operations is a missed opportunity for Security Studies and for the discipline's potential contribution to policy.

This paper seeks to address that gap by providing a conceptual exegesis of influence operations. In the first part of the paper, I engage in concept-building by constructing a definition of influence operations in component parts. 'Influence operations' is not a new term, but practitioners have not historically focused on precisely defining the term or its boundaries; meanwhile, contemporary studies focusing on the subject implicitly disagree over how to define influence operations and what is included in the category of behavior. By clearly defining influence operations, I show how it relates (but is different than) other terms such as information operations, disinformation, and political warfare, and why it is analytically useful. In the second section of the paper, I argue that influence operations are under-studied within International Relations and Security Studies, likely due to the historic sidelining of public opinion and Intelligence Studies within the discipline. In the third section of the paper, I argue that influence operations are prevalent, and are likely to remain so, for at least three reasons: they are cost-effective, due to the rising costs of war and declining costs of reaching foreign publics; they are difficult to deter, because of features of the information environment; and foreign propagandists are likely to exaggerate their impact, giving reason to continue to wage operations *even if* single operations don't have strong first-order effects. Fourth, I contend that studying influence operations can strengthen the International Security Studies discipline by expanding notions of state power beyond the traditional bean-count measures and challenging traditional models of

[26] Yadav, Kamya. 2020. "Countering Influence Operations: A Review of Policy Proposals Since 2016." *Carnegie Endowment for International Peace Report*.
[27] In the Appendix, I show the results of a "study of studies"—I find that few articles in top journals in Political Science, International Relations, and Security Studies include the term "influence operation", "information operation", "political warfare", "psychological warfare" or "psychological operations" at least twice.

foreign policy decision-making that treat the domestic and foreign as distinct boards. Fifth, I conclude by charting directions for future research on influence operations, motivating the subsequent two papers of this dissertation.

Conceptualizing Influence Operations

Despite the renewed interest in the policy community on non-kinetic forms of statecraft, influence operations are "elusive to define" and no agreed-upon definition exists.[28] CIA analyst Richards J. Heuer, Jr. wrote in 1990 that "[t]here is no agreement as to what is encompassed by the term influence operations"; that has not changed in the years since.[29]

A field without clear definitions is problematic. It creates opportunity for people to study different things under the same name and then argue about contradicting conclusions. The literature on terrorism—where scholars disagree about whether to determine the universe of cases based on the action or the actor—has suffered from such definitional splits.[30] As Silove argues regarding the grand strategy literature, "The strength of theories or research programs depends on the soundness of the concepts on which they are based."[31] This is particularly important as researchers begin to create datasets that track and classify influence operations.[32] As Sartori writes, "concept formation stands prior to quantification."[33] By constructing a definition in component parts, I aim to make my decisions for what qualifies, and what does not, clear. I justify where my boundaries are by showing what types of operations or categories of behavior are included or excluded based on the criteria. Others may disagree with where I draw the bounds, but clearly delineating the bounds is the prerequisite for such a debate.

I define influence operations as "*covert informational efforts intended to affect attitudes or opinions in a manner favorable to the perpetrator*." This definition has three components

---

[28] Schneier, Bruce. "8 Ways to Stay Ahead of Influence Operations," *Foreign Policy,* August 12, 2019. See also: Pamment, James and Alicia Wanless. 2019. "How Do You Define a Problem Like Influence?" *Journal of Information Warfare* 18(3): 1-14.
[29] Heuer, Richards Jr. 1990. "Assessing Soviet Influence Operations." *The Works of Richard Heuer, Jr.*, pp. 3.
[30] Asal, Victor, Luis De La Calle, Michael Findley, and Joseph Young. 2012. "Killing Civilians or Holding Territory? How to Think about Terrorism." *International Studies Review* 14: 475-497.
[31] Silove, Nina. "Beyond the Buzzword: The Three Meanings of 'Grand Strategy.'" *Security Studies* 27(1): 27-57.
[32] E.g. *DFRLab/Dichotomies-of-Disinformation*, https://github.com/DFRLab/Dichotomies-of-Disinformation. See also: Martin, Diego A., Jacob N. Shapiro, and Julia Ilhardt. "Trends in Online Foreign Influence Efforts: Version 2.0." *Working Paper*.
[33] Sartori, Giovanni. 1970. "Concept Misformation in Comparative Politics." *The American Political Science Review* 64(4): 1033-1053.

worth unpacking: (1) covert, (2) informational efforts, and (3) intended to affect attitudes or opinions in a manner favorable to the perpetrator. In this section, I primarily use examples of state-backed political operations but, as I describe below, influence operations can also be used by other actors and toward other ends.

Influence Operations are Covert

The first criterion is the covert nature of influence operations. Influence operations are a type of covert action, operations which are planned and executed in a way that hides true sponsorship and that, if uncovered, are designed such that the sponsor can plausibly claim deniability.[34] If State A attempts to influence State B in a way that is entirely transparent—including the source of the information, the relationship between initiator and target, etc.—then the effort would not qualify as an influence operation. The criterion of covert distinguishes influence operations from other types of activities that are overt, like high-level visits for public diplomacy or white propaganda.

A burgeoning literature in International Relations studies covert action, but primarily addresses kinetic forms of covert action (like assassinations and covertly arming rebels).[35] As a result, research has not systematically theorized the different ways in which informational efforts can be covert. I posit that actors waging influence operations can conceal their involvement in at least four ways: *covert identity*, *covert sourcing*, *covert coordination*, and *covert sponsorship*.[36]

In the case of covert identity, an actor attempts to hide or obfuscate the true identity of the proximate person or entity conveying information.[37] During World War II, both Axis and Allied powers ran radio stations that claimed to be from the other side; for example, the British Political Warfare Executive ran Soldatensender Calais, which pretended it was German military broadcasting. In the social media age, foreign actors often run accounts that pretend to be other

[34] This is an adapted version of the US Government's definition of covert operations codified in NSC 10/2.
[35] For recent advances in the covert regime change literature, see: Carson, Austin. 2018. *Secret Wars: Covert Conflict in International Politics*. Princeton, New Jersey: Princeton University Press; Downes, Alexander B., and Mary Lauren Lilley. 2010. "Overt Peace, Covert War?: Covert Intervention and the Democratic Peace." *Security Studies* 19(2): 266–306; Joseph, Michael F., and Michael Poznansky. 2018. "Media Technology, Covert Action, and the Politics of Exposure." *Journal of Peace Research* 55(3): 320–335.
[36] These categories are not exhaustive but demonstrate a range of ways of concealing identity.
[37] Historically, black propaganda has been considered propaganda claiming to come from a source different from the true one and grey propaganda is unattributable propaganda.

people—such as when accounts operated from Iran pretended to be Americans and tweeted about U.S. politics during the 2020 election campaign.[38]

In covert sourcing, an actor attempts to hide how they obtained information they spread or to use false sources, even if the identity of the proximate communicator is transparent. This is often the case in narrative laundering operations.[39] Brownfield describes how Moscow placed huge numbers of false stories in non-Communist press during the Cold War. Brownfield writes of his interview of Alezander Krasnechev, senior KGB office in Rangoon, Burma:

> His department was responsible for receiving drafts of articles from Moscow, translating them into Burmese, and then seeing that they were placed in local publications to appear as if they had been written by Burmese authors. The final step was to send copies back to Moscow. From there they were quoted in Soviet broadcasts of publications as evidence of "Burmese opinion" that favored the Communist line.[40]

In the social media era, state media and front organizations use the covert sourcing tactic when they knowingly imbed quotations or tweets in stories from fake personas. As DiResta and Grossman show, Russian military intelligence, the GRU, have cited many now-exposed fake personas in GRU-affiliated front outlets.[41] Unlike in covert identity, what is hidden in covert sourcing is not the true identity of the proximate person conveying information—such as the news publication—but that of their source(s).

In covert coordination, actors spreading information attempt to appear independent when they are, in fact, working in concert. Social media influence operations often use many accounts to share similar or identical content to give the false impression of popular support when the accounts are, in fact, coordinated. In December 2019, Twitter removed 88,000 coordinated accounts that were run by Smaat, a social media marketing company, and shared political content such as tweets critical of the governments of Qatar, Iran, and Turkey.[42] These accounts attempted

---

[38] Goldstein, Josh A. and Lindsey Hundley. 2021. "Viva la #Resistance? Now-Suspended Twitter Accounts from Iran Claim #Resistance Personas in US Politics and Middle Eastern Affairs." *Stanford Internet Observatory Report*.
[39] Shires, James. 2020. "The Simulation of Scandal: Hack-and-Leak Operations, the Gulf States, and U.S. Politics," *Texas National Security Review* 3(4): 10-29.
[40] Brownfield 1984, 6; as cited Jowett, Garth S. and Victoria O'Donnell. 2012. *Propaganda and Persuasion: Fifth Edition.* Sage Publications. p. 24.
[41] DiResta, Renée and Shelby Grossman. 2019. "Potemkin Pages & Personas: Assessing GRU Online Operations, 2014-2019." *Stanford Internet Observatory Report.* Accessed: https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/potemkin-pages-personas-sio-wp.pdf
[42] Twitter Safety. "New disclosure to our archive of state-backed information operations." *Twitter Blog*. December 20, 2019. Accessed: https://blog.twitter.com/en_us/topics/company/2019/new-disclosures-to-our-archive-of-state-backed-information-operations.html; DiResta, Renée, Shelby Grossman, K.H., and Carly Miller. 2019. "Analysis of

to appear as independent users, and an average social media user would lack the knowledge that they were operated by a single entity.

Finally, in covert sponsorship, actors hide or obfuscate the relationship between the proximate source of information and a sponsor guiding that information. This is the case when organizations publish or receive editorial direction from a source not transparent to readers. During the Cold War, Radio Free Europe broadcast commentary to approximately 35 million global listeners; the fact that the CIA funded the station was secret until the late 1960s.[43] A cyber-age example is the Strategic Culture Foundation, an online journal registered in Russia that, according to the U.S. Treasury, is directed by the SVR and closely affiliated with the Russian Ministry of Foreign Affairs. The Strategic Culture Foundation publishes Western fringe voices and conspiracy theorists and yet conceals its relationship to the Russian government.[44]

Table 1: Examples of 'Covertness' in Influence Operations

Types of Covert Behavior	**Description of Behavior**	**Example of Behavior**
Covert identity	Actors attempt to hide or obfuscate the identity of the person or entity conveying information	• Black or grey propaganda • Running agents of influence
Covert sourcing	Actors attempt to hide how they obtained information they spread or to covertly manipulate information, even if their identity is transparent	• Citing fake personas • Hack and leaks • Forging documents
Covert coordination	Actors attempt to hide their coordination and appear as independent voices	• Coordinated media campaigns
Covert sponsorship	Actors attempt to hide or obfuscate the relationship between the proximate source of information and a sponsor guiding their behavior	• Using front organizations to manipulate opinion • Planting news stories in legitimate press

Twitter Takedown of State-Backed Operation Attributed to Saudi Arabian Digital Marketing Firm Smaat." *Stanford Internet Observatory Report.*

[43] Johnson writes that nearly one-third of Soviet adults and almost one-half of East European adults were regular listeners, and more than $150 million annually was spent on efforts to jam Radio Free Europe broadcasts. Johnson, A. Ross. 2010. *Radio Free Europe and Radio Liberty: The CIA Years and Beyond.* Washington, DC: Wilson Center Press.

[44] United States Department of State. 2020. "Pillars of Russia's Disinformation and Propaganda System." *Global Engagement Center Special Report.* p. 15. Accessed: https://www.state.gov/wp-content/uploads/2020/08/Pillars-of-Russia%E2%80%99s-Disinformation-and-Propaganda-Ecosystem_08-04-20.pdf
The Strategic Culture Foundation front has also been used to smear specific political candidates, such as during Evelyn Farkas's run for Congress. See: Barnes, Julian E. "State Dept. Traces Russian Disinformation Links." *The New York Times*. August 5, 2020. Accessed: https://www.nytimes.com/2020/08/05/us/politics/state-department-russian-disinformation.html

These different types of covert informational behavior all meet the criterion of "covert" for influence operations. Furthermore, they can be deployed independently or in conjunction (e.g. when the identity and source are both covert). For example, in hack and leak operations an actor may first hide that they obtained the materials by hacking the victim (covert sourcing) and then hide their identity when releasing information (covert identity). This framework creates opportunity for researchers to study the relative effectiveness of different tactics of covert behavior, an area that Bateman et al. identify as lacking research.[45]

The covert criterion is also central to the significance of influence operations as a category of behavior. It raises the question: why do actors pursue the covert route in the informational domain? Two answers from the existing literature on covert action include (1) operational effectiveness, and (2) political repercussions.[46] In terms of operational effectiveness, actors benefit from hiding their identity, sources, coordination, or sponsorship. Kux writes, "To be effective, agents of influence must remain clandestine. Once exposed, they lose all utility."[47] Likewise, if a sockpuppet account (an account that pretends to be someone it is not) is exposed as such, it will be less credible as a cue-giver to its target audience. The notion that the identity of the source affects how information is perceived by a target is supported by a rich literature in Political Science and psychology on cue-taking.[48]

In addition to operational effectiveness, states may pursue covert influence operations to limit political repercussions. Work on kinetic forms of covert action suggests that covert action is a tool to limit escalation, to limit political repercussions, and to avoid public perceptions of breaking domestic and international law.[49] To date, these justifications have not been thoroughly explored in the non-kinetic influence operations literature, but they provide a prima facie

[45] Bateman et al. write that no study in their review "directly tested for potential variations in effectiveness between different influence operations tactics." See: Bateman, Jon, Elonnai Hickock, Laura Courchesne, Isra Thange, and Jacob N. Shapiro. 2021. "Measuring the Effects of Influence Operations: Key Findings and Gaps From Empirical Research." *Carnegie Endowment for International Peace*. Accessed: https://carnegieendowment.org/2021/06/28/measuring-effects-of-influence-operations-key-findings-and-gaps-from-empirical-research-pub-84824

[46] Poznansky, Michael. 2020. "Revisiting plausible deniability." *Journal of Strategic Studies*.

[47] Kux, Dennis. 1985. "Soviet Active Measures and Disinformation: Overview and Assessment." *Parameters* 15(4).

[48] Hovland, Carl I. and Walter Weiss. 1951. "The Influence of Source Credibility on Communication Effectiveness." *Public Opinion Quarterly* 15: 635-650. Guisinger, Alexandra and Elizabeth N. Saunders. 2017. "Mapping the Boundaries of Elite Cues: How Elites Shape Mass Opinion Across International Issues." *International Studies Quarterly* 61(2): 425-441.

[49] Poznansky 2020.

expectation that states would pursue covert informational activity in part based on public perception.[50]

Influence Operations are Informational

Second is the specification of informational efforts. This helps to distinguish influence operations from kinetic, violent operations designed primarily to affect attitudes or opinions. For example, if State A attacks State B not to defeat soldiers or gain territory but to coerce State B's politicians, this would not qualify as an influence operation. Atomic attacks on Hiroshima and Nagasaki were designed to coerce the Japanese government—to undermine enemy morale. The primary target was not the population killed, but the minds of decision-makers.[51] The criterion of "informational" distinguishes influence operations from these violent actions.

Although violent efforts can be designed to influence a target in a desired manner, if any type of influence qualifies as an influence operation then the term suffers from conceptual stretching and loses its analytical utility.[52] The informational criteria distinguishes influence operations from covertly arming rebels, as the United States has done in Afghanistan and Syria, and from covert regime change.[53] However, this criterion does not preclude influence operations from inciting violence or amplifying violent narratives; in these cases, the first-order content of the operation is still informational. For example, Iranian persistent threat actors created a website "Enemies of the People" which listed images, home addresses, and other personal information of several U.S. officials in 2020. The website—while concealing that it was created by Iranian actors—called for violence against those on the list and superimposed crosshairs over their images.[54] This qualifies as an influence operation because the actor concealed its identity and spread information designed to affect opinions in a particular manner.

---

[50] They also motivate by subsequent two dissertation papers, which study a mechanism of operational effectiveness (Paper 2) and whether plausibly deniability in fact limits support for retaliation, among other outcomes (Paper 3).
[51] Sánchez-Cuenca, Ignacio and Luis de la Calle. 2009. "Domestic Terrorism: The Hidden Side of Political Violence." *Annual Review of Political Science* (12)1: 31-49.
[52] Sartori 1970.
[53] Mazzetti, Mark, Adam Goldman, and Michael S. Schmidt. "Behind the Sudden Death of a $1 Billion Secret C.I.A. War in Syria," *The New York Times*, August 2, 2017. Accessed: https://www.nytimes.com/2017/08/02/world/middleeast/cia-syria-rebel-arm-train-trump.html
Pattison, James. 2015."The Ethics of Arming Rebels." *Ethics & International Affairs* 29(4): 455–471.
[54] Nakashima, Ellen, Amy Gardner, and Aaron C. Davis. "FBI links Iran to online hit list targeting top officials who've refuted Trump's election fraud claims." *The Washington Post.* December 22, 2020. Accessed: https://www.washingtonpost.com/national-security/iran-election-fraud-violence/2020/12/22/4a28e9ba-44a8-11eb-a277-49a6d1f9dff1_story.html

Influence Operations are Designed to Affect Attitudes/Opinions in a Manner Favorable to the Perpetrator

Third is the specification that influence operations are intended to affect attitudes or opinions in a manner favorable to the perpetrator. This suggests that influence operations have an intention and are not accidental. It also reflects that influence operations are conducted for diverse ends. They may be used to promote or smear a political actor, to shift currently held attitudes or reinforce those attitudes, to dispel false information circulating or inoculate a population to an informational campaign they believe to be pending.[55] The perpetrator does not need to influence the target to do something; they may also influence the target to do nothing.

The criteria laid out above apply to influence operations in multiple domains: covert, informational campaigns can be waged by prominent individuals, businesses, NGOs, governments, and other entities. For example, a recent study by the U.S. Federal Communications Commission (FCC) found that 8.5 million fake comments to the FCC over its proposal to scale back internet regulation were paid for by Broadband for America, an advocacy group funded by the nation's top internet providers.[56] Similarly, in 2019, the U.S. Federal Trade Commission reached a $2.5 million settlement with Devumi, a company that sold fake indicators of social media influence to companies and individuals alike.[57] These operations meet the criteria of an influence operation: the perpetrator attempted to hide their identity by using fake accounts, the goal was to affect opinion in a targeted way, and the means were informational, not kinetic. Defining influence operations as a type of activity—rather than by the perpetrator or the goal—opens opportunity for Security Studies scholars and political scientists to learn lessons from covert, informational campaigns in other areas. This is significant because the very marketing

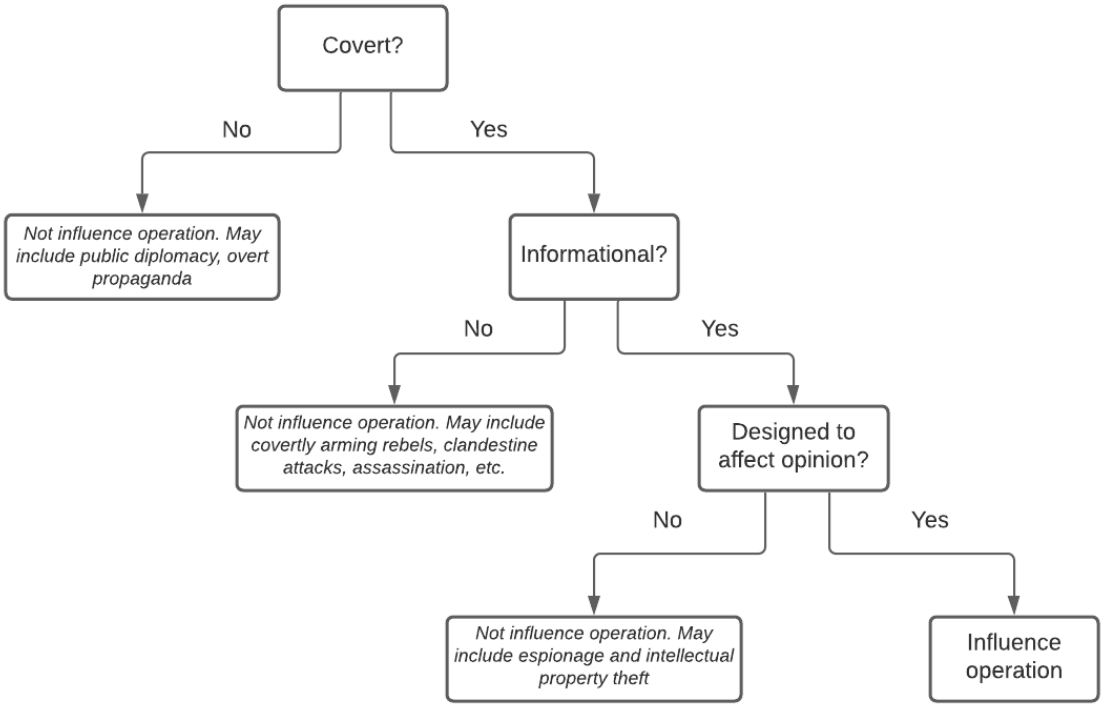
[55] For examples of many such operations, see United States Department of State. 1986. *Active Measures: Process of Anti-U.S. Disinformation and Propaganda Campaigns*. Department of State Publication 9630. Accessed: http://insidethecoldwar.org/sites/default/files/documents/Soviet%20Active%20Measures%20Substance%20and%20Process%20of%20Anti-US%20Disinformation%20August%201986.pdf

[56] Grimaldi, James V. "FCC's Net Neutrality Proposal Marred by Millions of Fake Comments." *The Wall Street Journal.* May 6, 2021. Accessed: https://www.wsj.com/articles/fccs-net-neutrality-proposal-marred-by-millions-of-fake-comments-11620309351

[57] Federal Trade Commission. "Devumi, Owner and CEO Settle FTC Charges They Sold Fake Indicators of Social Media Influence; Cosmetics Firm Sunday Riley, CEO Settle FTC Charges That Employees Posted Fake Online Reviews at CEO's Direction." Federal Trade Commission Press Release. October 21, 2019. Accessed: https://www.ftc.gov/news-events/press-releases/2019/10/devumi-owner-ceo-settle-ftc-charges-they-sold-fake-indicators

and PR firms that are increasingly implicated in foreign political influence operations also apply those tactics for non-state, apolitical clients.[58]

Figure 1: Key Criteria Distinguishing Influence Operations from Other Types of Behavior



Foreign Political Influence Operations

For this paper, I focus on a subset of influence operations: those that are waged for political ends and that target people in a different country. This subset of influence operations is important for International Security Studies, which examines how states seek to protect themselves and advance their interests in an anarchic world.[59]

[58] Grossman, Shelby and Khadeja Ramali. 2020. "Outsourcing Disinformation." *Lawfare*. Accessed: https://www.lawfareblog.com/outsourcing-disinformation; Silverman, Craig, Jane Lytvynenko, and William Kung. "Disinformation For Hire: How a New Breed of PR Firms Is Selling Lies Online." January 6, 2020. *BuzzFeed News*. Accessed: https://www.buzzfeednews.com/article/craigsilverman/disinformation-for-hire-black-pr-firms; Gleicher, Nathaniel, Margarita Franklin, David Agranovich, Ben Nimmo, Olga Belogolova, and Mike Torrey. 2021. "The State of Influence Operations 2017-2020" *Facebook Report*. Accessed: https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf

[59] Waltz, Kenneth N. 1959. *Man, the State, and War: A Theoretical Analysis*. Topical Studies in International Relations; New York.

In this paper and dissertation, I refer to "influence operations" as a shorthand for "foreign political influence operations." But for setting a research agenda, it is important to establish first that influence operations need not be foreign. Actors can wage domestic information operations: the Royal Thai Army, for example, launched an influence operation using fake Twitter accounts that promoted the Army and Prime Minister Prayut Chan-o-Cha to domestic audiences;[60] Harouna Douamba used front news organizations on Facebook to demean France and launch a disinformation campaign claiming that the local UN mission in the Central African Republic was engaged in arms trafficking;[61] actors in Saudi Arabia have repeatedly used domestic trolls to distort the domestic online information environment.[62] Defining influence operations to include domestic and foreign behavior creates room for future scholarship to assess how leaders learn from their domestic experiences with information control and influence operations when they try to shape public opinion covertly abroad.

For the purpose of this paper, the criterion of foreign distinguishes foreign influence operations from domestic operations. Several clarifications are in order. First, a foreign influence operation by actors in State A does not need to target all of State B. The size and nature of an operation may range from targeting the general public, to a subset of elites, to specific ethnic groups, to a particular decision-maker. Second, foreign influence operations do not have to target citizens of State B: they can target diaspora groups or potential defectors of State A who are in State B at the time of operation, for example. Third, foreign influence operations can capitalize on, inspire, or work in coordination with domestic overt or covert informational efforts. Foreign actors can amplify content that is already in circulation, and domestic actors may wittingly or unwittingly spread content from a foreign influence operation.[63]

The criterion of political distinguishes political influence operations from topics that are unrelated to political outcomes. Covert, informational campaigns that target businesses may have

---

[60] Goldstein, Josh A., Aim Sinpeng, Daniel Bush, Ross Ewald, and Jennifer John. 2020. "Cheerleading Without Fans: A Low-Impact Domestic Information Operation by the Royal Thai Army." *Stanford Internet Observatory Report*.

[61] Hundley, Lindsay, Renee DiResta, Josh A. Goldstein, Shelby Grossman, Cooper Reed, Adriana Stephan, and Julia Thomas. 2021. "One Face, Many Names: An Investigation into Fake NGOs and Media Outlets Linked to Harouna Douamba on and off Facebook." *Stanford Internet Observatory Report*.

[62] Barrie, Christopher and Alexandra A. Siegel. 2021. "Kingdom of Trolls? Influence Operations in the Saudi Twittersphere." *Journal of Quantitative Description: Digital Media* 1: 1-41.

[63] See Starbird, Kate, Ahmer Arif, and Tom Wilson. 2019. "Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations." *Proceedings of the ACM on Human-Computer Interaction* 3: 1-26.

economic, not political, motives. For example, a covert, informational campaign—using a false story—claimed that Coca-Cola issued a recall for Dasani water because the water was contaminated with a parasite.[64] Such an effort could qualify as an influence operation if it is covert and designed to affect attitudes or opinions of a target in a manner favorable to the perpetrator. However, such operations focused on hurting businesses fall outside of my subset of interest: political influence operations.

Assessing the Utility of My Concept and Definition

Influence operations is a term that has been used by intelligence practitioners and policymakers, but has not been clearly defined such that there is an established consensus definition.[65] Two principles are particularly important for defining a concept that has been in use: first, does the definition capture as much as possible of the politically relevant phenomena without conceptual stretching? Clearly delineating the concept helps us capture as much of possible of the phenomenon that has discoverable boundaries—i.e. that exists in a self-evident way in the real-world. Examples of a concept should 'belong' together. I show that my definition captures covert, informational behaviors that intelligence practitioners have historically considered to be influence operations. I further show that my definition includes a broad reading of different components of an influence operation (the perpetrator, target, aim, and content) compared to several other definitions in a way that better captures empirical examples that obviously belong.

Second, is the concept sufficiently differentiated from other concepts that are more widespread to warrant its use? Influence operations is not a new concept, but to justify my use of it—and my argument that more scholars should focus on this concept—I need to establish that it does not simply capture the same (or very similar) behaviors as other adjacent topics. To do this, I show how my three criteria of an influence operation: (1) covert, (2) informational, and (3) intended to affect attitudes or opinions in a manner favorable to the perpetrator are what Gerring calls *sometimes-differentiating* criteria; they distinguish influence operations from related

[64] Associated Press. "AP Fact Check: Coca-Cola not recalling Dasani water." *Associated Press*. January 24, 2017. Accessed: https://apnews.com/article/ecfb398fa555456d99fe40e0a05dd1da; See also Ferraro, Matthew F. and Jason C. Chipman. "Fake news threatens our businesses, not just our politics." *The Washington Post.* February 8, 2019. Accessed: https://www.washingtonpost.com/outlook/fake-news-threatens-our-businesses-not-just-our-politics/2019/02/08/f669b62c-2b1f-11e9-984d-9b8fba003e81story.html

[65] Pamment and Wanless 2019.

phenomena. I showed examples of this differentiation in Figure 1, but further elaborate how influence operations is a term that is distinct from other terms used in the field such as disinformation, information operations, and political or psychological warfare.

Capturing Relevant Phenomena

As Gerring writes in his foundational work on concept formation, there should be coherence in grouping among phenomena captured by a concept. "Covert, informational efforts" provides the core of the definition, creating coherence in the group. Influence operations are efforts to influence opinions or attitudes, where actors opt for secrecy, not transparency, and use information, not violence. This creates a common motivation behind different types of influence operations—secrecy for operational effectiveness and/or political repercussions, as described above.

By clearly delineating the boundaries of the concept, I aim to capture as much as possible of the politically relevant phenomenon. The definition above includes the types of operations that have frequently been called influence operations by practitioners and scholars of intelligence.[66] These operations include:

- Spreading propaganda that is attributed to a false source (black propaganda) or that is unattributed (grey propaganda)
- Forging documents
- Planting news stories in legitimate press
- Using front organizations to influence opinion abroad
- Covertly spreading rumors
- Leaking stolen materials
- Running agents of influence to covertly shape opinion abroad

Within the existing research on influence operations, many of the definitions fail to capture some of the politically relevant phenomenon listed above. When we construct boundaries for a concept that already exists as a term, we know if a category is too narrow if it excludes empirical examples that obviously belong within the boundaries. This is the case with definitions of

[66] Sometimes these operations have also been called "active measures," but active measures can also include covert action which is not informational but kinetic or violent—such as assassinations and sabotage operations. Godson and Schultz describe many of these operations as "covert political techniques." Godson, Roy and Richard Schultz. 1985. "Soviet active measures: Distinctions and definitions." *Defense Analysis* 1(2): 101-110. See also: United States Department of State. 1981. "Soviet Active Measures: Forgery, Disinformation, Political Operations." Special Report No. 88. Accessed: https://www.hsdl.org/?abstract&did=807615

influence operations that define components of an influence operation narrowly in terms of the perpetrator, aim, content, or target of the operation.

First, scholars disagree over whether influence operations are defined by the perpetrator. Some scholars define influence operations as a tool used by authoritarians. Kalathil writes that "digital influence operations," for the purpose of his study, "will be used to broadly capture the categories of digital activity most commonly employed by authoritarian regimes internationally to manipulate, censor, and degrade the integrity of the information space for strategic purposes."[67] Others, such as Weedon et al., take a more general approach, defining influence operations as "[a]ctions taken by governments or organized non-state actors to distort domestic or foreign political sentiment."[68]

Defining influence operations based on the identity of the perpetrator as authoritarians is too narrow because it fails to capture that democracies wage influence operations as well. As recent studies of active measures have shown, the United States has a long history of influencing attitudes abroad while concealing its hand.[69] During the Cold War, the United States covertly planted falsehoods in foreign press about carcinogenic Soviet spy dust, Soviet sponsorship of international terrorism, and attempts by Bulgarians to assassinate the pope.[70] More recently, Facebook removed fake accounts linked to "individuals associated with French military" that commented on politics in the Central African Republic (CAR) and attempted to combat a Russian influence operation targeting CAR audiences.[71] It is not clear why these instances—from the United States and actors in France—should not count as influence operations, but the same actions by an nondemocratic state should. My definition, like Shapiro et al., Nimmo, and

[67] Kalathil, Shanthi. 2020. "The Evolution of Authoritarian Digital Influence: Grappling with the New Normal." *PRISM* 9(1): 34.
[68] Weedon, Jen, William Nuland, and Alex Stamos. 2017. "Information Operations and Facebook." *Facebook*. Accessed:https://i2.res.24o.it/pdf2010/Editrice/ILSOLE24ORE/ILSOLE24ORE/Online/_Oggetti_Embedded/Documenti/2017/04/28/facebook-and-information-operations-v1.pdf Note that Weedon et al. use the terms information operations and influence operations interchangeably.
[69] Rid 2020; Shimer, David. 2020. *Rigged: America, Russia and 100 Years of Covert Electoral Interference* HarperCollins Publishers.
[70] Alexandre 1988, 114-115; as cited in Jowett and O'Donnell 2012, pp. 25-26.
[71] Graphika and the Stanford Internet Observatory. 2020. "More-Troll Kombat: French and Russian Influence Operations Go Head to Head Targeting Audiences in Africa." *Joint Technical Report by Graphika and the Stanford Internet Observatory*. Accessed: https://public-assets.graphika.com/reports/graphika_stanford_report_more_troll_kombat.pdf

Weedon et al., does not specify the nature of the target state or the relationship between perpetrator and target, thus capturing covert, informational campaigns waged by diverse actors.[72]

Second, there is disagreement over the target of an influence operation. Some definitions suggest that influence operations, by definition, target specific categories of actors. Pamment et al. write, "When we talk about information influence activities, we are talking about the systematic use of deceptive techniques to undermine democracy"—implying that democracies are the target of influence operations.[73] The RAND corporation writes that the goal of an operation is "competitive advantage over an opponent." Meanwhile, other definitions such as Martin et al. do not specify attributes of a target state, simply specifying the target as "another state."[74]

Narrow definitions of the target of an influence or the relationship between the perpetrator and the target fail to capture covert informational campaigns that target allies or that target non-democracies. In practice, states can, and do, covertly manipulate attitudes in countries they do not consider to be opponents, as the United Kingdom did in the United States prior to Pearl Harbor or as the United States did in Italy in the lead-up to the 1948 election. O'Donnell and Jowett correctly note that "even allies target friendly nations with black propaganda" (a type of influence operation).[75]

Third, there is disagreement over the aim of an operation. Some scholars provide definitions which include a broad array of potential aims. Larson et al. write that influence operations aim to "foster attitudes, behaviors, or decisions by foreign target audiences that further U.S. interests and objectives."[76] The Facebook security team likewise defined influence operations as aiming "to achieve a strategic and/or geopolitical outcome."[77] These definitions take a value-neutral approach to the aim. By contrast, other definitions specify harmful aims.

---

[72] Nimmo, Ben. 2020. "The Breakout Scale: Measuring the Impact of Influence Operations," *Brookings Report*. Accessed: https://www.brookings.edu/research/the-breakout-scale-measuring-the-impact-of-influence-operations/
[73] Pamment, James, Howard Nothhaft, and Alicia Fjällhed. 2019. "Countering Information Influence Activities: A Handbook for Communicators." *Swedish Civil Contingencies Agency Report*.
[74] Martin, Shapiro, and Ilhardt 2020.
[75] Jowett and O'Donnell 2012, p. 19.
[76] Larson, Eric V. Richard E. Darilek, Daniel Gibran, Brian Nichiporuk, Amy Richardson, Lowell H. Schwartz, and Cathryn Quantic Thurston. 2009. *Foundations of Effective Influence Operations: A Framework for Enhancing Army Capabilities*, RAND Corporation Monograph Series.
[77] Weedon et al, "Information Operations and Facebook."

Jeangène-Vilmer et al. define the activity based on "the political intention to cause harm" and Pamment et al. specify a desire to "create a climate of distrust between a state and its citizens."[78]

My definition, like Larson et al., does not specify the aim of an operation beyond affecting attitudes or opinions in an intended manner, because covert informational campaigns are waged for diverse ends.[79] States can covertly influence opinion without a desire to cause harm. Members of the British Special Operations Executive would not have thought that attempts to encourage American citizens to favor intervention against Nazi Germany was causing harm. CIA operatives countering a communist-backed candidate in Italy did not intend to cause harm. And Soviet influence measures designed to strengthen the European peace movement described in the introduction were probably not considered by the USSR as causing harm. If we only include cases where actors intend to cause harm, we fail to understand the range of aims for which covert informational campaigns are deployed, and are prone to bias our case selection based on culture-specific conceptions of harm.

Fourth, there is disagreement over the content. Some definitions suggest that the content of an influence operation may be "true or false"[80] while others suggest it is "false information or information that is consciously distorted" (Jeangène-Vilmer et al.). Martin et al. are even more specific, defining influence operations as "producing content designed to appear indigenous to the target state."[81]

I side with the Defending Digital Democracy project in including true or false information. A 1954 paper prepared by the "Operations Coordinating Board"—the group that oversaw U.S. covert operations under President Eisenhower—outlined coordination of "gray activities." These are activities in which "[t]he true source (U.S. Government) is not revealed to the target audience. The activity engaged in plausibly appears to emanate from a non-official American source, or an indigenous, non-hostile source, or there may be no attribution."[82] Scholars who define influence operations based on veracity of the information discount gray

[78] Pamment et al. 2019.
[79] Larson et al. 2009.
[80] Cederberg, Gabriel, Jordan D'Amato, Corinna Fehst, Simon Jones, Kunal Kothari, Aleksandra Milcheva, and Irene Solaiman. 2019. "National Counter-Information Operations Strategy." *Belfer Center Report.* Accessed: https://www.belfercenter.org/publication/national-counter-information-operations-strategy.
[81] Martin, Shapiro, and Ilhardt, "Trends in Online Foreign Influence Efforts."
[82] Gray, Gordon and Elmer B. Staats. "Foreign Relations of the United States, 1950–1955, The Intelligence Community, 1950–1955 - Office of the Historian," accessed November 11, 2020, https://history.state.gov/historicaldocuments/frus1950-55Intel/d181

activities. This is problematic because it contradicts how practitioners have conceived of influence operations—often claiming that influence operations will be more effective if they rely on the truth.[83] Scholars who specify that content must appear indigenous exclude false flag operations, which are also historically common.

Each of these disagreements can be classified based on whether the author(s) take a broad or narrow conception of influence operations, as summarized in Table 2. Together, they show dissensus in the existing literature over what qualifies—and what does not qualify—as an influence operation. My definition of influence operations sides with a broad definition of the term; it defines influence operations as an action, not a type of information or an action that can only be taken by a particular perpetrator. (For a survey of existing definitions of influence operations, see Table 3.) It then narrows to foreign political influence operations as a subset that is of particular interest to International Security Studies.

Table 2: Summary of Definitional Disagreements: Broad Versus Narrow Constructions of Interstate Influence Operations Across Four Dimensions

	My Definition	**Narrow Criteria Found in the Literature**
Perpetrator	• Any actor	• States • Authoritarians
Target	• Any actor	• An opponent • A democracy
Aim	• Foster attitudes/behaviors that are favorable to the perpetrator	• Cause harm • Foster distrust • Undermine democracy
Content	• True or false information (i.e. not specified)	• False information • Designed to appear indigenous to the target

[83] Pacepa, Mihai and Ronald J. Rychlak, 2013. *Disinformation: Former Spy Chief Reveals Secret Strategies for Undermining Freedom, Attacking Religion, and Promoting Terrorism.* WND Books.

Table 3: Survey of existing definitions of influence operations

Martin and Shapiro 2019	FIEs [foreign influence operations] are defined as: (i) coordinated campaigns by one state to impact one or more specific aspects of politics in another state, (ii) through media channels, including social media, by (iii) producing content designed to appear indigenous to the target state.
RAND corporation (site page)	Information operations and warfare, also known as influence operations, includes the collection of tactical information about an adversary as well as the dissemination of propaganda in pursuit of a competitive advantage over an opponent.
Jeangène-Vilmer et al. 2018 (Ministry of Foreign Affairs – Ministry for the Armed Forces, France)	"Information manipulation" encompass[es] three criteria: a coordinated campaign, the diffusion of false information or information that is consciously distorted, and the political intention to cause harm.
Defending Digital Democracy's "National Counter-Information Operations Strategy" 2019	Information operations, also referred to as influence operations, are the dissemination of information, true or false, that seeks to manipulate public opinion and/or influence behavior. The motivation for conducting information operations ranges from political to social to economic, with state and non-state actors leveraging technological and psychological tools to achieve their goals.
Larson et al. 2009 (followed by A. Ross Johnson 2018 and others)	Influence operations are the coordinated, integrated, and synchronized application of national diplomatic, informational, military, economic, and other capabilities in peacetime, crisis, conflict, and postconflict to foster attitudes, behaviors, or decisions by foreign target audiences that further U.S. interests and objectives.
Pamment et al. 2018	Information influence activities involve potentially harmful forms of communication orchestrated by foreign state actors or their representatives. They constitute deliberate interference in a country's internal affairs to create a climate of distrust between a state and its citizens. Information influence activities are used to further the interests of a foreign power through the exploitation of perceived vulnerabilities in society. …When we talk about information influence activities, we are talking about the systematic use of deceptive techniques to undermine democracy.
Weedon et al. 2017 (Facebook)	Information (or Influence) Operations - Actions taken by governments or organized non-state actors to distort domestic or foreign political sentiment, most frequently to achieve a strategic and/or geopolitical outcome. These operations can use a combination of methods, such as false news, disinformation, or networks of fake accounts (false amplifiers) aimed at manipulating public opinion.
Christopher Wray 2019 (Director of the FBI)	Foreign influence operations—which include covert actions by foreign governments to influence U.S. political sentiment or public discourse…
Nimmo 2020	"Influence operations" are defined as efforts to influence public or political debate and decisionmaking processes that rely in part or in whole on covert activity.

Concept Differentiation

In addition to showing that the boundaries constructed in my definition above successfully capture relevant political phenomenon, I must also show that the term is sufficiently different from other terminology that may be more widespread to justify its use. Gerring writes that "if all one is doing is rearranging the same set of parts into (essentially) the same whole—with a new label—then not much has been accomplished."[84] 'Influence operations' is not a new label, but more well-known terminology in the field exists. Influence operations, which captures covert, informational campaigns, is distinct from this existing terminology in the Security Studies literature—including information operations, disinformation, and political/psychological warfare. Many of these other terms have been used in different ways, giving them ambiguous meaning.

Some use information operations and influence operations interchangeably; this is problematic because, within military academic and practitioner circles, an information operation is defined as "the integrated employment, during military operations, of IRCs in concert with other lines of operation to influence, disrupt, corrupt, or usurp the decision making of adversaries and potential adversaries while protecting our own."[85] A significant portion of these activities—such as defending one's own information capabilities—operate on logics that are unrelated to covertly shaping opinion abroad. Furthermore, some information operations—like cyber-attacks to degrade information systems—do not operate through information. Influence operations as defined in this paper is thus a term that is distinct from information operations, and captures a meaningful category of behavior.

Another common term related to information manipulation is disinformation. Disinformation is conventionally used to refer to "false or misleading information spread with the intent to deceive." In this definition, "intent to deceive" distinguishes disinformation from the more common misinformation.[86] Influence operations, in contrast to disinformation, captures a

[84] Gerring, John. 1999. "What Makes a Concept Good? A Criteria Framework for Understanding Concept Formation in the Social Science." *Polity* 31(3): 383.
[85] See: Joint Publication 3-13 "Information Operations" 2014, ix. Accessed: https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_13.pdf
[86] Brooking, Emerson T. and Suzanne Kianpour. 2020. "Iranian Digital Influence Efforts: Guerilla Broadcasting for the Twenty-First Century." *Atlantic Council Report*. Accessed: https://www.atlanticcouncil.org/wp-content/uploads/2020/02/IRAN-DIGITAL.pdf

category of behavior, not a type of information. Influence operations can be true or false, but they are covert. Disinformation can be covert or overt, but it must be false. Thus, influence operations and disinformation are conceptually distinct but *can* overlap: a covert informational campaign (an influence operation) can spread false information with the intent to deceive (disinformation).

More recently, some scholars have begun to use disinformation to refer to content spread with the intent to deceive. This is problematic for two reasons: first, it is often difficult to establish that someone's intent is to deceive; second, moving the definition away from its more widespread use (of false information) creates contradictions with existing literature and renders the term potentially ambiguous. Influence operations thus more clearly captures covert, informational campaigns.

Finally, there are a number of terms in the Security Studies literature that have been used to capture attempts to influence public opinion in adversarial countries. This includes psychological warfare and political warfare. Psychological warfare is a term that is often defined based on its use within the context of war or active hostilities,[87] and includes activities ranging from providing aid to deploying terrorism. Political warfare likewise includes a broader range of activities than influence operations; as George Kennan writes, "political warfare is the employment of all the means at a nation's command, short of war, to achieve its national objectives. Such operations are both overt and covert." Kennan includes actions such as political alliances, economic aid, alongside propaganda in his definition of political warfare.[88] These behaviors are much more varied and operate on different logics. (In the Appendix, I show how influence operations are different from, and overlap with, other related terms on each of the criteria in my definition.)

Jack, Caroline. 2017. "Lexicon of Lies: Terms for Problematic Information." *Data & Society Report*. Accessed: https://datasociety.net/pubs/oh/DataAndSociety_LexiconofLies.pdf; Wardle, Claire and Hossein Derakhsahn. 2018. "Thinking about 'information disorder': formats of misinformation, disinformation and mal-information." In Ireton, C., & Posetti, J. (Eds.) Journalism, 'fake news', & disinformation: handbook for journalism education and training. UNESCO. pp. 43-53.

[87] E.g. see: Crossman, R. H. S. 1952. "Psychological Warfare." *Royal United Services Institution Journal* 98(592): 528-539; Schleifer, Ron. 2006. "Psychological operations: A new variation on an age old art: Hezbollah versus Israel." *Studies in Conflict & Terrorism* 29(1): 1-19.

[88] Kennan, George F. Policy Planning Staff Memorandum 269, Washington, D.C.: U.S. State Department, May 4, 1948. Accessed: http://academic.brooklyn.cuny.edu/history/johnson/65ciafounding3.htm as cited by Robinson, Linda, Todd C. Helmus, Raphael S. Cohen, Alireza Nader, Andrew Radin, Madeline Magnuson, and Katya Migacheva. 2019. "The Growing Need to Focus on Modern Political Warfare." *RAND Report*.

Why Influence Operations are Under-Explored in International Security Studies Scholarship

In the section above, I provided an overview of influence operations, breaking down the term in component parts and showing how influence operations differ from other common terminology in the field. Despite the historical importance and contemporary prominence of influence operations, they are not well-studied in top journals of Political Science or International Relations. As I report in Table 5 in the Appendix, among eight top journals in Political Science, International Relations, and International Security Studies, only one article uses the term 'influence operation(s)' twice. Adjacent terms such as information operations and political warfare are likewise not prominent.

In this section, I posit that the lack of focus on influence operations may be a product of the development of the International Security Studies and International Relations fields vis-à-vis public opinion and intelligence. Influence operations are designed to affect attitudes or behaviors, often of mass publics. Thus, literature on public opinion formation would be a natural home for studying the effects of an operation. However, within the International Relations discipline, public opinion literature has only recently begun to consider foreign cues and influence operations are a relatively under-explored subject. A second likely angle for research might be studying the perpetrators of influence operations; historically, they have been intelligence organizations. However, here, too, there is limited work in Political Science and International Security Studies, given that Intelligence Studies has historically been sidelined from mainstream International Relations work.

Public Opinion and International Relations Scholarship

In his seminal study on the evolution of International Security Studies, Barry Buzan begins "International Security Studies (ISS) grew out of debates over how to protect the state against external and internal threats after the Second World War.[89] In the immediate aftermath of World War II, no single question dominated the field. Questions of political warfare—including

[89] Buzan, Barry and Lene Hansen. 2009. "Defining International Security Studies" in *The Evolution of International Security Studies*. Cambridge University Press. p. 8. Note that many associate the birth of International Relations with the founding of the first Chair for the study of international politics at University College of Wales, Aberystwyth in 1919. See: Schmidt, Brian C. "On the history and historiography of international relations." In: Carlsnaes W., Risse T. and Simmons BA (eds) *Handbook of International Relations.* Sage. pp. 3-28.

influence operations under a different name—received some consideration.[90] But by 1955 to 1965, "the sub-field became almost exclusively devoted to the study of nuclear weapons and bipolar rivalry."[91] Theorists were concerned with nuclear deterrence, polarity, and issues related to escalation. As the Cold War developed, International Security Studies scholars retained their rationalist approaches. Some scholars went beyond the black box of the state, considering the role of bureaucratic politics on decision-making or the role of psychology and elite misperceptions.[92] But, during this time period, neither intelligence nor public opinion was a primary focus.

Research on public opinion, however, ballooned after the Vietnam War. Research during the first two decades after World War II produced the Almond-Lippmann consensus, which Holsti describes as the belief that public opinion is volatile, irrational, and that "in the final analysis, [public opinion] has little if any impact on foreign policy."[93] Research on public opinion during the Vietnam War undermined this position and produced a new conventional wisdom—that the public is not volatile, is rational, and responds to elite cues and to events. More recently, scholars have suggested the public opinion does play a role in foreign policy decision-making: publics can place audience costs[94] and elites are more likely to make decisions that are favored by their domestic publics.[95] Recent studies address public support for the use of force against cyber-attacks and public support for using nuclear weapons.[96]

Despite the increased focus on public opinion in International Relations, influence operations remain outside the major focus. This may be due to their foreign source. Scholars

---

[90] E.g. Qualter, Terence H. 1962. *Propaganda and Psychological Warfare*. Studies in Political Science. New York: Random House.

[91] Buzan and Hansen 2013, p. 67.

[92] See Allison, Graham T. 1971. *Essence of Decision: Explaining the Cuban Missile Crisis*. Boston: Little, Brown and Company. Jervis, Robert. 1976. *Perception and Misperception in International Politics*. Princeton University Press.

[93] Holsti, Ole R. 1992. "Public Opinion and Foreign Policy: Challenges to the Almond-Lippmann Consensus Mershon Series: Research Programs and Debates." *International Studies Quarterly* 36(4): 439.

[94] Weeks, Jessica L. P. 2008. "Autocratic Audience Costs: Regime Type and Signaling Resolve." *International Organization* 62(1): 35-64.

[95] Baum, Matthew A. 2004. "How Public Opinion Constrains the Use of Force: The Case of Operation Restore Hope." *Presidential Studies Quarterly* 34(2): 187-226; Tomz, Michael, Jessica L.P. Weeks, and Keren Yarhi-Milo. 2020. "Public Opinion and Decisions About Military Force in Democracies." *International Organization* 74(1): 119-143.

[96] Kreps, Sarah and Jacquelyn Schneider. 2019. "Escalation Firebreaks in the Cyber, Conventional, and Nuclear Domains: Moving beyond Effects-Based Logics." *Journal of Cybersecurity* 5(1); Sagan, Scott D. and Benjamin A. Valentino. 2017. "Revisiting Hiroshima in Iran: What Americans Really Think about Using Nuclear Weapons and Killing Noncombatants." *International Security* 42(1): 41-79.

have historically argued that foreign sources are unlikely to be seen as credible in the eyes of domestic publics, and thus foreign cues are ineffective.[97] However, applying this reasoning to influence operations fails on both empirical and theoretical grounds. Empirically, recent work has shown that foreign political cues—state visits, statements from foreign leaders—can have an effect on public opinion.[98] Experimentally manipulating whether American respondents presented with a *Russia Today* article were told of the article's source, Fisher finds that foreign cues from unfriendly states can have a persuasive effect.[99] More importantly, this type of foreign political cue is distinguishable from influence operations: it is public diplomacy, which does not involve the concealment necessary to qualify as an influence operation. The central objection to the relevance of public diplomacy—suspicion towards foreign elites and sources—may not apply to influence operations. Influence operations are covert, and thus the identity of the perpetrator can be hidden from the public eye. For example, forgeries, journalists for hire, and other forms of influence operations can appear—or attempt to appear—as indigenous to the target state. This therefore will not suffer backlash from being from a foreign source which may be thought to be non-credible. While there are opportunities for International Security scholars to explore the effect of influence operations on public opinion, limited research to date has taken this direction.[100]

The Sidelining of Intelligence Studies

Part of the reason influence operations are under-explored or under-theorized may be due to the sidelining of intelligence. As referenced above, influence operations have historically fallen within the domain of intelligence agencies. Yet, as of the 1990s, Der Derian wrote that

---

[97] Mermin, Jonathan. 1999. *Debating War and Peace: Media Coverage of US Intervention in the Post-Vietnam Era*. Princeton University Press.

[98] Goldsmith, Benjamin E. and Yusaku Horiuchi. 2011. "Spinning the Globe? U.S. Public Diplomacy and Foreign Public Opinion." *The Journal of Politics* 71(3): 863-875; Hayes, Danny and Matt Guardino. 2011. "The Influence of Foreign Voices on U.S. Public Opinion," *American Journal of Political Science* 55(4): 831-851; Goldsmith, Benjamin E. and Yusaki Horiuchi. 2012. "In Search of Soft Power: Does Foreign Public Opinion Matter for US Foreign Policy?" *World Politics* 64(3): 555-585; Agadjanian, Alexander and Yusaku Horiuchi. 2020. "Has Trump Damaged the U.S. Image Abroad? Decomposing the Effects of Policy Messages on Foreign Public Opinion." *Political Behavior* 42(2): 581–602.

[99] Fisher, Aleksandr. 2020. "Demonizing the Enemy: The Influence of Russian State-Sponsored Media on American Audiences." *Post-Soviet Affairs*. 36(4): 281-296.

[100] There have been exceptions in political science or other adjacent fields such as communication studies. For example, see Bail, Christopher A., Brian Guay, Emily Maloney, Aidan Combs, D. Sunshine Hillygus, Friedolin Merhout, Deen Freelon, and Alexander Volfovsky. 2020. "Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017." *PNAS* 117(1): 243-250.

Intelligence Studies is "the least understood and most undertheorized area of international relations."[101]

Amy Zegart argues that academics have been suspiciously absent from the study of intelligence for two primary reasons. First, Zegart argues that Intelligence Studies sits in a "theoretical no-man's-land" between American politics and International Relations. International Relations scholars cover relationships between states, not within them, and American politics scholars focus on domestic regulatory agencies but not on other bureaucracies.[102] Second, Zegart notes that intelligence agencies are particularly difficult to study from the outside due to classification of information, inaccessibility of technically unclassified information, and difficulties in finding the right people to interview.[103] "Few foreign-policy options within the federal government are held as closely as covert action, and no intelligence operations—with the exception of domestic spying—have been so controversial."[104]

Other reasons suggested include that academic historians have been "repelled by the evident absurdity of many best-selling publications on the subject";[105] aversion to studying a subject seen to be, at least to a degree, distasteful;[106] and the belief among some liberal scholars that the subject is associated with political conservatism.[107] As a result, Walton finds that "when it comes to historical studies of foreign policies and International Relations, intelligence is either wholly missing or treated as an aberration, even in the best and most recently published accounts."[108] In 2018, he writes that while a full-text search on *JStor* raises 29,303 results containing the words "U.S. foreign policy," when combined with "National Security Agency" "the results plummet to just 303 listed results-approximately one percent of the larger total."[109]

---

[101] As cited in Herman, Michael. 1996. *Intelligence Power in Peace and War.* Cambridge: Cambridge University Press. p. 3.
[102] Zegart, Amy B. 1999. *Flawed by Design: The Evolution of the CIA, JCS, and NSC*. Stanford University Press; Amy Zegart. 2007. "Cloaks, Daggers, and Ivory Towers: Why Academics Do Not Study U.S. Intelligence" *Strategic Intelligence*. Westport, CT: Praeger.
[103] Zegart 2007.
[104] Johnson, Loch. 1989. "Covert Action and Accountability: Decision-Making for America's Secret Foreign Policy." *International Studies Quarterly* 33(1): 82.
[105] Andrew Christopher and David Dilks. 1984. *The Missing Dimension: Governments and Intelligence Communities in the Twentieth Century*. Basingstoke: Macmillan.
[106] Jervis, Robert. 2018. "Intelligence and International Politics," in *The Oxford Handbook of International Security* by Robert Jervis, Alexandra Gheciu and William C. Wohlforth, ed. Oxford University Press. pp. 515-530.
[107] Loch Johnson citing emails with Rose McDermott 2012.
[108] Walton, Calder. 2018. "Historical Amnesia: British and U.S. Intelligence, Past and Present," *Secrecy and Society* 2(1): 2.
[109] Ibid, p. 8

Work within Intelligence Studies draws on declassified documents to tell untold portions of international history or to analyze influence operations. For example, studies address German efforts to bolster Anglophobia in France in 1914-1915, U.S. propaganda activities in India during the 1950s, U.S. dissemination of anti-Communist and pro-democratic propaganda in Asia, British covert propaganda in the 1960s, British black broadcasting during the Suez crisis of 1956, U.S. mailings of articles and books to Eastern Europeans behind the Iron Curtain, Soviet *dezinformatsiya* and the activities of the bloc services, and lessons from influence operations through Radio Free Europe/Radio Liberty.[110] This list is illustrative (not representative or comprehensive) of the kinds of work on influence operations growing in this literature.

While studies in intelligence have begun to address untold components of history, the research area is relatively weak in its development of theory. Michael Warner writes that scholars of intelligence are split between those who conceive of intelligence as information for decisionmakers and those who see intelligence as warfare by other means. "Without a clear sense of the dependent variable," he continues, "we find it difficult to understand which independent factors cause and affect intelligence phenomena."[111] Johnson, in opening the monumental "Strategic Intelligence," a 5-volume collection of essays on intelligence, notes that "[t]heories have been difficult to construct in the intelligence field because the empirical data base for constructing and testing theories still remains relatively thin."[112] However, as Rid shows in his groundbreaking work on active measures, it is possible to rigorously study the activities of intelligence agencies within history.[113] We also have a significant number of resources from the

---

[110] See, respectively: Wilkin, Bernard and Maude Williams. 2017. "German Wartime Anglophobic Propaganda in France, 1914–1945." *War in History* 24(1): 28–43; Pullin, Eric D. 2011. "'Money Does Not Make Any Difference to the Opinions That We Hold': India, the CIA, and the Congress for Cultural Freedom, 1951–58." *Intelligence and National Security* 26(2-3): 377-398; Lombardo, Johannes R. 1999. "A Mission of Espionage, Intelligence and Psychological Operations: The American Consulate in Hong Kong, 1949–64." *Intelligence and National Security* 14(4): 67; Cormac, Rory. 2019. "Techniques of Covert Propaganda: The British Approach in the Mid-1960s." *Intelligence and National Security* 34(7): 1064-1069; Rawnsley, Gary D. 1996. "Overt and Covert: The Voice of Britain and Black Radio Broadcasting in the Suez Crisis, 1956." *Intelligence and National Security* 11(3): 497-522; Matthews, John P. C. 2003. "The West's Secret Marshall Plan for the Mind." *International Journal of Intelligence and Counterintelligence* 16(3): 409-427; Holland, Max. 2006. "The Propagation and Power of Communist Security Services *Dezinformatsiya*." *International Journal of Intelligence and Counterintelligence* 19(1): 1-31; Johnson, A. Ross. "Managing Media Influence Operations: Lessons from Radio Free Europe/Radio Liberty." *International Journal of Intelligence and Counterintelligence.* 31(4): 681-701; Warner, Michael. 2009. "Intelligence Theory: Key Questions and Debates," in Mark Phythian, Peter Gill, and Stephen Marrin, ed., *Intelligence Theory: Key Questions and Debates*. Studies in Intelligence Series. London: Routledge. pp. 16-17.

[111] Warner 2009.

[112] Johnson, Loch. 2007. *Strategic Intelligence*, Volume 2: Intelligence and the Quest for Security. Westport, Conn.: Praeger Security International. p. 3.

[113] Rid 2020.

information security community looking at current influence operations on social media campaigns. Twitter, for instance, has created a public archive of state-backed influence operations removed from their platform. This creates opportunity for political scientists to contribute theory and theory-testing on the subject of influence operations.[114]

Why Security Studies Should Consider Influence Operations

In this section, I argue that influence operations are important to study because they matter to the contemporary security landscape and they will continue to matter in the future. Influence operations are prevalent in the current security environment for three reasons: First, they are low cost. Influence operations in the social media era are cheap, particularly compared to potential kinetic alternatives. Second, propagandists are likely to inflate the first-order impact of operations and exaggerate their success. This may be conscious—they have bureaucratic incentives to attract resources and mandates—but it also may be subconscious—because the cognitive bias of overconfidence (or positive illusions) is likely to thrive in domains where disconfirming evidence is rare. Propagandists also may inflate their success because targets are likely to over-react. Together, these two reasons—low cost and over-confidence on the part of propagandists—provide a prima facie reason for U.S. adversaries to wage influence operations: all else equal, before the United States responds, they explain why adversaries would be inclined to wage these operations.

I further argue that influence operations are likely to remain relevant in the near to medium term because they are difficult to deter. Part of the U.S. failure to respond effectively to influence operations in recent years has been partisan and political. But even if the United States put its best foot forward, I argue that it would still be difficult to deter influence operations. Following classic deterrence theory, I distinguish between deterrence-by-denial and deterrence-by-punishment. For the former, deterrence-by-denial is difficult because there are many vulnerable points, many ways to reach target audiences, and control of these nodes is distributed across different sectors. Deterrence-by-punishment is difficult because of attribution challenges and because U.S. policymakers fear escalation.

[114] For a model of this work, see Barrie and Siegel 2021.

Low Cost of Influence Operations

Security studies scholars generally recognize a foundational principle of realism—that states are in an international system without a global policeman to ensure order. Under conditions of anarchy, states must provide for their own security. *How* states provide for their own security, however, is not determined by anarchy alone. When costs of warfare are high and potential benefits are low, we should expect non-kinetic tools of statecraft—like influence operations—to become relatively more prominent. Today, we see these conditions bear out.

The rising prominence of influence operations is a product of two converging trends—first, the rising costs of war, and second, the declining costs of reaching foreign publics. While the rising costs of war occurred earlier in the 20th century, the declining costs of reaching foreign public has been a more recent development with the dispersion of internet access and social media. In this section, I briefly trace the trends in turn.

Rising Costs of War

Andrew Krepinevich traces a series of military "revolutions" from the Hundred Years' War through the late Cold War that transformed how war is fought over time.[115] Successive introductions of new technologies and operational codes—from the use of railroads and telegraphs which allowed states to more efficiently mobilize troops and sustain combat in World War I, to "combustion engine, aircraft design, and radio and radar" that allowed for the *blitzkrieg* of World War II—increased the potential destructiveness of war.[116] Tracing the loss exchange ratio of battles based on a time variable, Rosen and Horowitz find that more recent battles exhibit lower kill ratios.[117] Thus, both sides are more likely to endure significant costs of fighting.

---

[115] Krepinevich, Andrew F. 1994. "Cavalry to Computer: The Pattern of Military Revolutions." *The National Interest*. Accessed: https://nationalinterest.org/article/cavalry-to-computer-the-pattern-of-military-revolutions-848?page=0%2C2. Here, I accept Krepinevich's argument detailing introductions of new technologies and codes which have impacted the conduct of war. One can accept the general descriptive argument regarding changes in warfare without accepting all changes described as Revolutions in Military Affairs (RMAs).

[116] Krepinevich 1994; Porter 1994, as cited by Levy, Jack S. and William R. Thompson. 2011. *The Arc of War: Origins, Escalation, and Transformation*. Chicago, Ill.: University of Chicago Press. p. 142.
These changes in warfare have sometimes been attributable to military research and development (like submarine warfare), whereas at other times they have been driven by new civilian technologies (like the railroad). See: Cohen, Eliot A. 1996. "A Revolution in Warfare." *Foreign Affairs* 75(2): 37-54.

[117] Horowitz, Michael and Stephen Rosen. 2005. "Evolution or Revolution?" *Journal of Strategic Studies* 28(3): 445.

In addition to the rising costs of war, the potential benefits of war have declined. For millennia prior to the Industrial Revolution, agriculture was the dominant economic activity and there was a close association between land and power. In such a time, war produced a clear benefit: "control over territory—additional land and the associated labor force—that added directly to both economic and political power."[118] Nineteenth-century changes greatly diminished the cost of war: "the integration of all or nearly all of the population into the unified society of the nation-state meant that the mere acquisition of territory did not by itself convey effective control of the resources sited on it, especially the all-important human resources."[119] These trends continued after World War II: Van Evera notes that "the shift toward knowledge-based forms of production in advanced industrial economies since 1945 has reduced the ability of conquerors to extract resources from conquered territories."[120] This diminishes the risk of war because conquest doesn't pay as well. Van Evera highlights a paradox in modern control of foreign territory: states would have to adopt harsh police measures to control newly-acquired territory, yet those very harsh police measures would undermine the benefit of controlling that new territory by diminishing economic output. The rising costs and declining benefits of war make traditional conflict a relatively unattractive tool; instead, states are prone to use other forms of statecraft to further their security.[121]

Declining Costs of Reaching Foreign (Liberal) Publics

Not only are the costs of war high and the potential benefits low—suggesting that states may turn to other tools for interstate competition—but the costs of using influence operations specifically are low. In *Lie Machines*, Philip Howard notes that influence operations in the past often required significant resources—in Operation Neptune, the Stasi placed fake Nazi era records at the bottom of a lake in Czechoslovakia. "In 1964, the records were then 'found' and shared at a dramatic press conference, thereby ending the careers of some West German politicians and impending the prosecution of war criminals."[122] As Fly et al. write, "These

[118] Kaysen, Carl. 1990. "Is War Obsolete?" *International Security* 14(4): 49.
[119] Ibid, p. 53.
[120] Van Evera, Stephen. 1990. "Primed for Peace: Europe after the Cold War." *International Security* 15(3): 14.
[121] This does not suggest that kinetic war will never be used by states or never benefit states. Rather, it is a comment on aggregate trends.
[122] Howard, Philip N. 2020. *Lie Machines: How to Save Democracy from Troll Armies, Deceitful Robots, Junk News Operations, and Political Operatives*. New Haven: Yale University Press. p. XV.

tactics, however, were often costly and time consuming with limited reach, in stark contrast to the ease with which technology now facilitates remote manipulation and low-cost individual targeting of any American with a smart phone and a social media account."[123]

Two trends have contributed to the declining cost of running influence operations: First, is the so-called "democratization of the news"—anyone can publish news or opinion at no or low cost, driving up the number of voices weighing in on politics. This ecosystem allows would-be propagandists to create faux news sites,[124] amplify hyper-partisan content, and even attempt to hire unwitting journalists to write for their sites.[125] The second trend driving the declining costs of influence operations is widespread Internet and social media penetration. The creation of social media giants Facebook and Twitter in 2004 and 2006 respectively has given direct access to mass publics in foreign countries that do not effectively block their use. This is disproportionately countries with liberal values, such as open access to information.[126] Influence operations run by governments, state-affiliates, or disinformants-for-hire can reach these populations easily.

States can hire workers on the cheap to distort online social media discourse. In 2013, the Turkish AKP recruited 6,000 workers to "a new social media team, known as the New Turkey Digital Office, which was responsible for converting AKP sentiments into trending hashtags."[127] The Chinese government has "long been suspected of hiring as many as 2 million people to surreptitiously insert huge numbers of pseudonymous and other deceptive writings into the stream of real social media posts, as if they were the genuine opinions of ordinary people" in

[123] Fly, Jamie, Laura Rosenberger, and David Salvo. 2018. "Policy Blueprint for Countering Authoritarian Interference in Democracies." *Alliance for Securing Democracy Report.* p. 10. Accessed: https://www.gmfus.org/publications/asd-policy-blueprint-countering-authoritarian-interference-democracies.
[124] DiResta and Grossman 2019.
[125] Stubbs, Jack. "Exclusive-Russian operation masqueraded as right-wing news site to target U.S. voters." *Reuters*, October 1, 2020. Accessed: https://www.reuters.com/article/usa-election-russia-disinformation/exclusive-russian-operation-masqueraded-as-right-wing-news-site-to-target-u-s-voters-sources-idINL1N2GS0FY
[126] Walker, Christopher. 2018. "What Is 'Sharp Power'?" *Journal of Democracy* 29(3): 11; DiResta, Renée. 2018. "Open Hearing on Foreign Influence Operations' Use of Social Media Platforms (Third Party Expert Witnesses)," p. 17 § Senate Select Committee on Intelligence. Accessed: https://www.justsecurity.org/wp-content/uploads/2019/07/Open-Hearing-on-Foreign-Influence-Operations%E2%80%99-Use-of-Social-Media-Platforms-Third-Party-Expert-Witnesses.pdf.
[127] Shearlaw, Maeve. "Turkish journalists face abuse and threats online as trolls step up attacks," *The Guardian*, November 1, 2016. Accessed: https://www.theguardian.com/world/2016/nov/01/turkish-journalists-face-abuse-threats-online-trolls-attacks

China.[128] Linvill and Warren recently compared the costs of Russia's IRA efforts to conventional weaponry, noting that while the reported monthly budget of Russia's IRA was $1.25 million, a single Northrop Grumman B-21 bomber costs more than $550 million.[129] The low cost of influence operations has led analysts to describe them as an asymmetric form of political warfare that favor the weak over the strong.

Second, a budding industry of disinformants-for-hire allows governments without know-how to outsource propaganda cheaply. A recent Stanford Internet Observatory report highlights a number of operations removed from Facebook and Twitter for coordinated inauthentic behavior that were run, in full or in part, by marketing firms. These firms include Estraterra, DotDev, aRep Global, New Waves, Newave, Flexell, Smaat, Archimedes Group, and CLS Strategies.[130] Outsourcing allows governments to forego start-up costs of hiring and training propagandists in the tactics of contemporary social media and the inefficiencies of government bureaucracies. Instead, it allows governments to rely on market pricing and benefit from plausible deniability. In one experiment, Jigsaw, a corporate think tank affiliated with Google, put the Russian disinformation marketplace to the test. They hired SEO Tweet, a firm found on the Russian marketplace, to create a disinformation campaign for a dummy website—costing only $250.[131]

Likely Overconfidence of Propagandists

In the section above, I argued that influence operations are a relatively attractive tool because of their low costs and the ease with which states can now reach foreign publics. This assumes that influence operations have some effect—if not, a low-cost operation that achieves no effect would still waste resources. In this section, I advance a different but complementary argument: features of the environment and decision-maker make state-backed propagandists

---

[128] King, Gary, Jennifer Pan, and Margaret E. Roberts. 2017. "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument." *American Political Science Review* 111(3): 484-501.

[129] Linvill, Darren and Patrick Warren. "Foreign Influence Operations Don't Need to Succeed to Be Effective," October 23, 2020. *Lawfare*. Accessed: https://www.lawfareblog.com/foreign-influence-operations-dont-need-succeed-be-effective.

[130] Cryst, Elena, Esteban Ponce de León, and Daniel Suárez Perez. 2020. "Bolivarian Factions: Facebook Takes Down Inauthentic Assets." *Stanford Internet Observatory Report*.
See also: Bradshaw, Samantha and Philip N. Howard. 2019. "The Global Disinformation Disorder: 2019 Global Inventory of Organised Social Media Manipulation." *Oxford Internet Institute Report*.

[131] Barrett, Paul M. 2019. "Disinformation and the 2020 Election: How the Social Media Industry Should Prepare" *NYU Stern Center for Business and Human Rights Report*. Accessed: https://issuu.com/nyusterncenterforbusinessandhumanri/docs/nyu_election_2020_report/1

likely to inflate their success. Following the literature on bounded rationality, I assume that propagandists "are goal-oriented and adaptive, but because of human cognitive and emotional architecture, they sometimes fail, occasionally in important decisions."[132] I argue that their inflation of success is partly due to the genuine perception of actors waging influence operations, and partly due to bureaucratic politics. Therefore, even actors who wage relatively ineffective operations will continue to do so, degrading trust and the information environment more broadly.[133]

In his historical study of active measures and disinformation campaigns, Rid touches on the tendency of intelligence agencies to inflate their success. He writes:

> It's not that analysts simply believed their own lies; it's that operators, driven by professional education as well as bureaucratic logic, tended to overstate rather than understate the value of their own disinformation work. Analysts would write after-action reviews and project memos that justified their efforts in terms that were clearer and more convincing than what had happened on the ground, where cause and effect remained entangled by design—exacerbating existing fissures and cracks, or tapping into existing grievances, or enhancing existing activism—all of which meant that engineered effects were very difficult to isolate from organic developments."[134]

In this section, building back from Rid's conclusion, I unpack some of the reasons why propagandists are inclined to inflate their success. First, it is difficult to measure the effectiveness of an operation because of features of the operational environment. This makes the cognitive bias of overconfidence particularly likely, since it is a domain without clear disconfirming feedback. Second, bureaucratic politics posits that people have a tendency to solve problems with the capabilities they have, to rely on their favored dominant indicators, and to inflate their bureaucratic responsibility. Third, effectiveness—and perceived effectiveness—depends on the response of the target. When targets exaggerate effects, they increase the perpetrator's confidence and increase the likelihood of further operations. Together, these arguments suggest that propagandists who have waged an operation on social media are likely to perceive their effect as significant. In turn, influence operations are likely to retain salience in the near term.

---

[132] Jones, Bryan D. 1999. "Bounded Rationality." *Annual Review of Political Science* 2(1): 297.
[133] For an application of the misperceptions literature to intelligence analysts, see Heuer, Richards J. 1981. "Strategic Deception and Counterdeception: A Cognitive Process Approach." *International Studies Quarterly* 25(2): 294-327.
[134] Rid 2020, p. 430.

When Measurement is Difficult, Overconfidence Thrives

The first reason why propagandists might be inclined to inflate their success is because overconfidence thrives in areas where disconfirming evidence is rare. Johnson writes that "people tend to: (1) overestimate their capabilities; (2) overestimate their control over events; and (3) underestimate their vulnerability to risk."[135] A significant literature documents that people are over-confident in domains as varied as their own sensitivity, their ability to estimate uncertain quantities, and their ability to pick stocks.[136] Within the International Security literature, overconfidence, or "positive illusions," helps explain why states go to war when rational models would suggest that parties should be able to compromise to avoid costs of war.[137] It may also provide an explanation for why propagandists are likely to continue to wage operations, *even if influence operations had limited first-order effects.*

Overconfidence is particularly likely in domains that lack clear feedback that would disconfirm one's perceptions of their own abilities. It is difficult to measure the effects of influence campaigns, which may allow propagandists to exaggerate their impact.[138] Wallenius and Nillson conduct a review of the scientific literature and find that "while much has been written about PSYOP, empirical studies of psychological effects [are] rare. Neither is there an overarching applicable method of how to measure such effects."[139] As Carrier-Sabourin notes, unlike in conventional military conflicts where success can be measure by lines on a map that indicate territory acquired or held and by asymmetrical body counts, success in the influence domain is "a multifaceted and convoluted enterprise."[140]

---

[135] Johnson, Dominic D. P. 2020. *Strategic Instincts: The Adaptive Advantages of Cognitive Biases in International Politics*. Princeton Studies in International History and Politics. Princeton, New Jersey: Princeton University Press. p. 51.

[136] See: Block, David and Richard A. Harper. 1991. "David Block and Richard A. Harper, "Overconfidence in Estimation: Testing the Anchoring-and-Adjustment Hypothesis." *Organizational Behavior and Human Decision Processes* 49: 188-207; David Block and Richard A. Harper, "Overconfidence in Estimation: Testing the Anchoring-and-Adjustment Hypothesis," *Organizational Behavior and Human Decision Processes*, Vol. 49 (1991), pp. 188–207; Kent, Daniel, David Hirshleifer, and Avanidhar Subrahmanyam. 1998. "Investor Psychology and Security Market Under- and Overreactions." *The Journal of Finance* 53(6): 1839-1885; Ames, Daniel R. and Lara K. Kammrath. 2004. "Mind-Reading and Metacognition: Narcissism, Not Actual Competence, Predicts Self-Estimated Ability. *Journal of Nonverbal Behavior* 28(3): 187-209.

[137] E.g. Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49(3): 379-414.

[138] Nimmo 2020, pp. 2-3.

[139] Wallenius, Claes and Sofia Nilsson. 2019. "A Lack of Effect Studies and of Effects: The Use of Strategic Communication in the Military Domain." *International Journal of Strategic Communication*. 13(5): 407.

[140] Carrier-Sabourin, Krystel. 2011. "Measuring Effects and Success in Influence Operations: Challenges, Limitations, and Opportunities," Contract Report, Defence R&D Canada. p. 1. Accessed: https://apps.dtic.mil/dtic/tr/fulltext/u2/a568381.pdf. See also: Matthews, Miriam, Christopher Paul, David Schulker,

There are at least three challenges to assessing the impact of an influence operation—even for those waging the operation. First is the issue of data collection. Propagandists may seek to measure attitudinal change or behavioral change, but it can be difficult to acquire this information—particularly if the operation targets citizens in enemy territory. Because influence operations are often covert, conducting polls for sentiment analysis risks exposing the operation. Second, even when data is available on the target's attitudes or behavior, it may be challenging to disentangle the effect of the influence campaign from other events or information sources.[141] Practitioners have often argued that propaganda should be deployed as a complement to material efforts. However, under these conditions, over-determination complicates measurement.[142] Third, and relatedly, is the challenge of distinguishing between genuine support versus coordinated foreign behavior. Because many operations acquire narratives that already exist and work to amplify those narratives, it is difficult to disentangle the effect of the operation from authentic promotion among domestic audiences. This is evident from work that looks at how Russians intermixed with genuine social media users in 2016.[143]

When data collection is challenging and the effect of influence efforts is difficult to disentangle from other events or genuine domestic activity, there is a high level of uncertainty around effectiveness. Without clear disconfirming evidence, propagandists are prone to exaggerate their impact.

Bureaucratic Politics and Dominant Indicators

In their seminal work on the bureaucratic politics model, Allison and Halperin describe foreign policy decisions not as the rational calculation of the state as a unitary actor, but as the product of the pulling and hauling of competing bureaucratic interests.[144] Bureaucracies have an incentive to inflate the importance of the tools they can provide to a country's overall foreign policy, potentially bringing funding to their coffers and centering their place at the decision-

---

and David Stebbins. 2020. "Frameworks for Assessing USEUCOM Efforts to Inform, Influence, and Persuade." *RAND Report*.

[141] Matthews et al. 2020, p. 2.

[142] Wallenius and Nilsson 2019, p. 408.

[143] Starbird, Kate. "The Surprising Nuance Behind the Russian Troll Strategy." *Medium*. October 20, 2018. Accessed: https://medium.com/s/story/the-trolls-within-how-russian-information-operations-infiltrated-online-communities-691fb969b9e4

[144] Allison, Graham T. and Morton H. Halperin. 1972. "Bureaucratic Politics: A Paradigm and Some Policy Implications." *World Politics*. 24: 40-79.

making table. As Gartner describes in his work on strategic assessment in war, organizations have interests such as "autonomy, budget, prestige, and essence (the image that an organization has of itself)."[145] To assess their performance, they rely on indicators—"quantitative, time-specific measures of complex information."[146] While Gartner is specifically focused on war, the same analysis is true of broader efforts of intelligence or other agencies waging influence operations.

Like other areas of the governmental bureaucracy, those engaged in propaganda and foreign influence operations are likely to "choose indicators that they expect to suggest that their implemented policy is successful."[147] As Rid writes, "Like all bureaucracies, secret organizations crave metrics and data, to demonstrate how well they perform in the never-ending governmental competition for resources."[148] Propagandists are likely to focus on engagement—the number of people reached by a message—and not impact—the number of people who change their attitude or behavior as a result of the campaign. This is because it is easier to measure engagement than impact. This is prone to producing over-estimation of effect, since engagement is a lower threshold than changing opinions or attitudes.

Playing on the Partisans

The third reason why propagandists are likely to inflate their impact is because effectiveness—and perceived effectiveness—depends on the response of the target. If an influence operation is partially or wholly uncovered, the response of the target is likely to be used as a metric for success (or failure) of the operation. If the target exaggerates the effect, they increase the perpetrator's confidence and increase the likelihood of further operations.

In the United States, the polarized nature of politics is likely to drive exaggeration of the effect of influence operations by the aggrieved party. Influence operations are employed in service of political aims. When their message builds on existing internal divisions, they can become highly politicized. Those whose interests align with the foreign actor are less likely to perceive the operation as a serious threat than those whose interests do not align with the effort. Polling after Russian interference in the 2016 election showed a significant partisan split. A

[145] Gartner, Scott Sigmund. 1999. *Strategic Assessment in War*. New Haven; London: Yale University Press. p. 29.
[146] Ibid, p. 43.
[147] Ibid, p. 45.
[148] Rid 2020, p. 14.

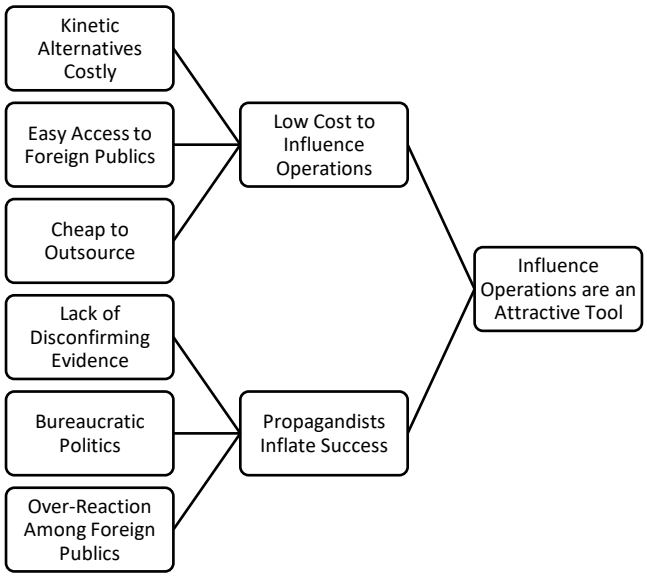
Suffolk Poll asked: "U.S. intelligence agencies concluded that Russia tried to meddle in the U.S. presidential election last year. How serious of an issue do you think this is - very serious, somewhat serious, not very serious, not at all serious?" While 81% of Democrats responded "Very Serious," only 24% of Republicans said the same.

Tomz and Weeks ran survey experiments in which they manipulated the type of foreign electoral intervention as well as the partisan identity of the candidate that the foreign country supported. They write:

> Instead of rejecting foreign involvement *tout court*, Americans exhibited a partisan double standard. Both Democrats and Republicans were more likely to condemn foreign involvement, lose faith in democracy, and call for retaliation when a foreign power sided with the opposition, than when a foreign power aided their own party.[149]

The same is true of influence operations which are not primarily focused on elections. People are likely to see intervention on behalf of a movement or idea they disagree with as more dangerous. Given the highly polarized nature of American politics, for example, there is often a perceived winning and losing side. Influence operations gaining press attention in the target country can itself be a sign of success; as they amplify internal divisions and gain attention, they provide evidence of impact that can bolster the propagandists' domestic credibility.

Figure 2: Why Influence Operations Are Seemingly an Attractive Tool



[149] Tomz and Weeks 2020, 857. I build on this work further in Paper 3.

Thus far, I have provided two reasons why influence operations are prevalent and likely to remain so. First, they are a prima facie cost-effective strategy. Second, propagandists are likely to exaggerate their effect. In the section to follow, I argue that even if the United States tried to deter influence operations—if it were not halted by political divides—it would be difficult.

Deterring Influence Operations

Even if the United States did not fail to respond to influence operations because of partisan or political issues, it would still be difficult to deter influence operations. Deterrence, as defined by the U.S. Department of Defense Joint Publication 3-0, is the "prevention of an adversary's undesired action by the existence of a credible threat of unacceptable counteraction and/or belief that the cost of action outweighs the perceived benefits." Like many recent pieces on deterrence, I follow Snyder 1961 in distinguishing between two deterrence strategies: deterrence-by-punishment and deterrence-by-denial.[150] Deterrence-by-punishment occurs when the threat of penalties for an attack deters an adversary from taking such steps. This closely fits with the Latin root of the term, *deterre*, which means "to frighten from or away."[151] Deterrence-by-denial, on the other hand, occurs when states make it costly or unlikely for an attack to succeed, reducing the adversary's confidence in his success.

A brief analysis demonstrates that it is difficult to deter influence operations by denial and by punishment, and, thus, influence operations are unlikely to cease in the near to medium term.

Deterrence-by-Denial

Deterrence-by-denial strategies operate by increasing the costs of an attack and decreasing its potential benefit. Two components are required for an influence campaign: first, a

[150] Pijnenburg Muller, Lilly and Tim Stevens. 2017. "Upholding the NATO Cyber Pledge." *Norwegian Institute of International Affairs Policy Brief;* Mazarr, Michael J. 2018. "Understanding Deterrence." RAND Perspective. Accessed: https://www.jstor.org/stable/resrep19896; Sweijs, Tim and Samo Zilincik. 2019. "Deterrence by Democracies in a Hybrid Context." *Cross Domain Deterrence and Hybrid Conflict* Hague Centre for Strategic Studies. Accessed: https://www.jstor.org/stable/resrep24191.6; Kaminska, Monica. 2021. "Restraint under conditions of uncertainty: Why the United States tolerates cyberattacks." *Journal of Cybersecurity* 7(1): 1-15.
[151] Freedman, Lawrence. 1989. "General Deterrence and the Balance of Power." *Review of International Studies* 15(2): 200.

perpetrator must select its message; second, a perpetrator must disseminate the message. Deterrence-by-denial strategies in the domain of influence operations, then, would operate by increasing the difficulty of accessing information or by limiting its potential spread. I focus on each of these two routes in turn.

Step I: Information Acquisition

The message is the "what" of an influence campaign—the narrative the perpetrator propagates. Messages are directly related to the aim of an operation. For example, the message used in a "chaos-producing operation" designed to confuse and disrupt is likely to be different from the message used in a "propaganda operation" designed to persuade.[152] Sowing discord can be done through spreading controversial or polarizing narratives; policy change requires more specific persuasive or disparaging messages.

I argue that perpetrators of influence operations can create or adopt the message of the campaign. In message creation, a perpetrator creates the narrative of an influence campaign, not pulling from existing discourse directly. This is often in the case of forgeries, hack-and-leaks, or outright lies, when the message of the influence operation would not exist without the campaign. In the second category—message adoption—actors don't create the narrative of an influence operation from scratch but amplify existing discourse.

It is difficult to impede an actor from creating or adopting a narrative for an influence campaign because of the competitive nature of international politics and the multitude of ways by which democracy could weaken. While the information war as a battleground between democracies and autocracies may be overly simplistic, if leaders perceive state leadership and credibility as competitive, the faltering of democracies benefits authoritarians and the faltering of authoritarians benefits democracies. Based on this view, any narrative which promotes divisions within the United States may be worth amplifying.

Step II: Information Dissemination

Features of the information environment make it difficult to limit the dissemination of material. While this has positive effects in holding domestic officials to account, for example,

[152] Lin, Herbert and Jaclyn Kerr. 2019. "On Cyber-Enabled Information Warfare and Information Operations." In Oxford Handbook of Cybersecurity.

these vulnerabilities can also be weaponized against democratic states. As an illustration, I highlight how states can use multiple vectors to manipulate foreign public opinion. These include social media, faux news sites, state media, and traditional media.

Social Media and Influence Operations: Much attention in the policy sphere has been given to social media influence operations by state-backed or state-aligned actors in the last few years. To wage social media influence operations, states can use a variety of tools. They may use sockpuppets, or "fictional identiti[es] created to bolster some point of view."[153] One reason to use sockpuppet accounts is to increase persuasiveness—targets are more likely to be persuaded by a messenger that they perceive as likeable and credible. Sockpuppets allow users to pose as individuals who are more likely to persuade the target, such as members of their own community. States can also use political bots, or "automated social media accounts, often built to look and act like real people, in order to manipulate public opinion."[154]

In addition to waging social media influence operations with sockpuppets or political bots, states can leverage genuine social media influencers who are aligned with the state to spread a particular message. Social media influencers—including politicians—may share the influence operation's narrative to their large followings. Such operations rely on cover coordination or sponsorship instead of covert identity.

Faux News Sites and Influence Operations: A second tool of dissemination is faux news sites. In addition to posing as specific individuals, states often construct or work with front organizations—such as front news outlets—to transmit their message to the public. While this may take the form of websites (i.e. not on social media), these websites may be created so their content can be spread on social media while giving the appearance of news. For example, DiResta and Grossman analyzed the fake media outlet, Inside Syria Media Center, which was a news front run by the Russian GRU.[155] While most of the GRU content received extremely low engagement, some of the Inside Syria Media Center stories were republished on dozens of "aligned media properties within days of its creation."[156]

---

[153] Poole, Steven. "What's the difference between a troll and a sockpuppet?" *The Guardian.* February 23, 2018. Accessed: https://www.theguardian.com/books/2018/feb/23/troll-steven-poole-word-of-week

[154] Woolley, Samuel. 2020. "Bots and Computational Propaganda: Automation for Communication and Control," in Nathaniel Persily and Joshua Tucker, ed., *Social Media and Democracy: The State of the Field, Prospects for Reform*. Cambridge: Cambridge University Press. pp. 89–110.

[155] DiResta and Grossman 2019.

[156] Ibid, p. 26.

State Media and Influence Operations: State media have attracted large public followings in foreign countries. More than 6.5 million people follow the Facebook Page for *Russia Today* (RT), a Russian state-run media organization in English. Over 100 million people follow CGTN—China Global TV Network—a state-controlled Chinese international media organization. These large state-backed media organizations can be coopted by states for influence operations, as in the case of covert sourcing. States may use fake accounts on social media to give the false impression of public support for a policy, and then use state media outlets to cover that (albeit fake) groundswell of opinion.

State media outlets have demonstrated creativity in how they disseminate their messages to foreign audiences. For example, conservative leaning websites in the United States agreed to join a distribution network that would allow members to display content on other members' pages. RT managed to join the network, and therefore had stories displayed on other news sites such as RealClearPolitics or Daily Caller.[157] It is important to recognize that the fact that content comes from a state media outlet does not make it an influence operation in itself—rather, state media outlets using deceptive tactics described in Section I, such as covert sourcing and covert coordination, qualify.

Traditional Media and Influence Operations: States can also capitalize on news coverage from legitimate media organizations in target countries to disseminate the message of their influence campaign. Two types of operations typify this approach: First, intelligence agencies will attempt to mislead foreign journalists in order to promote their favored narrative. Second, states will use hack-and-leak operations to garner press attention from legitimate news organizations in the target country with material that favors their interests. As Shires writes, these operations are "widely seen as a severe threat to liberal democratic structures…"[158]

These potential information dissemination channels each represent vulnerabilities. It is difficult for states to deter influence operations by denial because it requires building defenses on each dissemination channel described above. These channels are not directly controlled by government, requiring the participation of third-parties, such as social media companies.

---

[157] Hagey, Keach, Emily Glazer, and Rob Barry. "How Russia Today Skirts High-Tech Blockade to Reach U.S. Readers." *The Wall Street Journal.* October 7, 2020. Accessed: https://www.wsj.com/articles/how-russia-today-skirts-high-tech-blockade-to-reach-u-s-readers-11602078094

[158] Shires 2020.

Moreover, even if existing channels are more tightly regulated, foreign governments have other channels such as deceiving elites and decision-makers directly.

Deterrence-by-Punishment

If deterrence-by-denial is a challenge, states could instead deter through threats of retaliation. As Libicki writes in his work on cyber deterrence, the theory of deterrence-by-punishment rests on a state being "ready and willing" to impose costs on those who attack them.[159] This, however, requires being able to make timely and accurate attributions, and having the political will to significantly punish (or credibly threaten to significantly punish) those who wage influence. In this section, I outline reasons why it is difficult for states to credibly threaten retaliation at a level sufficient to deter operations.

Difficulty in Making Attribution

Much like in the field of cyber-attacks, determining the perpetrator of an influence operation is often challenging.[160] While some influence operations conceal veracity of information but make the source of the message known, many operations—like grey or black propaganda, for example—conceal the perpetrator. This is particularly true of social media operations in which the perpetrator of an operation conceals his or her identity in order to increase the likelihood that their message will be persuasive. As Libicki writes, adding ambiguity widens the number of potential attacks because weaker actors can threaten stronger ones; "Countries can use [information warfare] elements to harass countries they cannot hope to touch in traditional kinetic ways…"[161]

A trend that exacerbates the attribution problem is the outsourcing of disinformation campaigns.[162] When states rely on third-party for-profit companies, social media analysts may be

[159] Libicki, Martin C. 2018. "Expectations of Cyber Deterrence." *Strategic Studies Quarterly* 12(4): 44.
[160] Kello describes five characteristics of cyber that contribute to the attribution problem, including: proliferation of cyberweapons; anonymity in cyberspace; difficulty in distinguishing between technical signals and human perpetrators; difficulty in obtaining forensic evidence; and the likelihood that evidence will not be prompt enough for retaliation (2013, 33). Kello, Lucas. 2013. "The Meaning of the Cyber Revolution: Perils to Theory and Statecraft." *International Security*. 38(2): 7-40. For further work on the attribution problem, see: Lindsay, Jon R. 2015. "Tipping the Scales: The Attribution Problem and the Feasibility of Deterrence against Cyberattack." *Journal of Cybersecurity* 1(1): 53-67; Rid, Thomas and Ben Buchanan. 2015. "Attributing Cyber Attacks." *Journal of Strategic Studies* 38(1-2): 4-37.
[161] Libicki, Martin C. 2017. "The Convergence of Information Warfare." *Strategic Studies Quarterly* 11(1): 49-65.
[162] Silverman et al. 2020.

able to trace the operation to the third party, but cannot conclusively prove who put that third party to the task. This helps explain the proliferation of studies which assign *proximate* attribution of disinformation campaigns to marketing firms, but cannot conclusively determine *ultimate* attribution to a state sponsor.[163]

When attribution *can* be made to a foreign adversary directly, governments and researchers risk compromising sources and methods should they seek to hold the foreign perpetrator accountable. To prove to the international community that a foreign adversary is behind an attack, the attacked would have to provide evidence. But providing this evidence may provide insight into how the operation was attributed. This creates a challenge: either fail to provide evidence sufficient to convince third parties that a foreign adversary was behind an attack, or provide the evidence and allow the adversary to learn about your sources and methods. If details are shown, adversaries can update their methodology for future operations, which will thus be harder to detect. One example of this practice is the recent move to disinformation campaigns utilizing deep fake photographs for profile pictures; disinformants updated their strategy of stealing people's profiles after realizing that researchers could reverse-image-search profile pictures to see if the picture existed elsewhere on the Internet.

The discussion of attribution revealing sources and methods demonstrates the cat-and-mouse nature of influence operations detection. As platforms, researchers, and governments improve their ability to pick up on current disinformation, disinformants update their efforts in order to bypass these trends. This is evident in reports that Russia has outsourced propaganda creation to individuals in other countries, in order to avoid detection.[164] Thus, (a) attribution is difficult, (b) attribution to proxies complicates the picture, and (c) *even if* attribution can be made, public attribution risks exposing sources and methods and helping propagandists improve their strategies.

---

[163] DiResta, Renée, Shelby Grossman, and Alexandra Siegel. "In-House vs. Outsourced Trolls: Digital Mercenaries Shape State Influence Strategies." *Working Paper.*

[164] Ward, Clarissa, Katie Polglase, Sebastian Shukla, Gianluca Mezzofiore, and Tim Liste. "Russian election meddling is back – via Ghana and Nigeria – in your feeds." *CNN.* April 11, 2020. Accessed: https://www.cnn.com/2020/03/12/world/russia-ghana-troll-farms-2020-ward/index.html

Fear of Retaliation

Some states may not threaten costly punishment out of fear of retaliation. The United States may not want to respond to influence operations that are relatively low level out of fear that foreign adversaries will escalate their influence activities or move behavior into other realms. Escalation in conflict dynamics has been thoroughly explored in the Security Studies literature, particularly by those studying the role of nuclear weapons on interstate behavior.[165] Leaders may fear retaliation and escalation even if they are not explicitly bargaining over territory and are involved in interactions short of traditional conflict.

Recent reporting on Russia's interference in the U.S. 2016 election suggests fear of retaliation may have been a driving force in limiting U.S. retaliatory measures. Shimer argues that the Obama Administration knew about Russian efforts to influence public opinion on social media, but chose not to retaliate out of fear that the Kremlin would, in turn, escalate their behavior.[166] The Obama Administration was aware that the Kremlin had breached American voting systems and feared that should the Obama Administration retaliate for social media influence efforts, the Russians might exploit these breaches to manipulate votes.

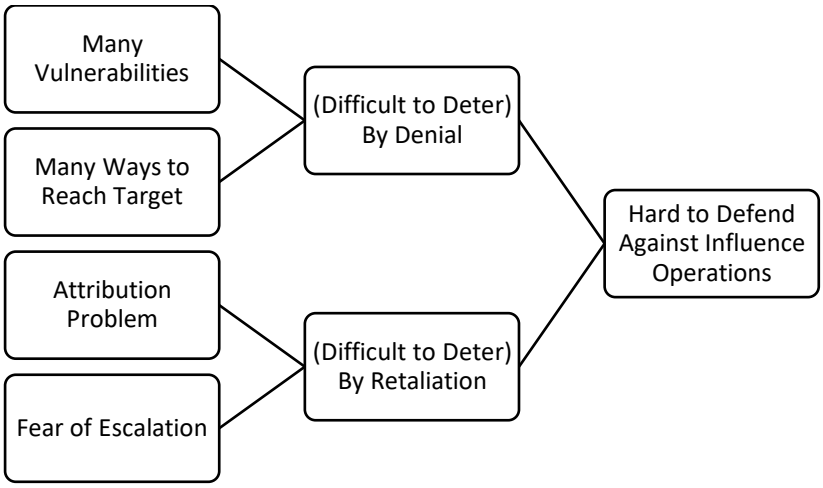
A similar logic has been articulated in Carson's work on covert regime change. Carson argues that states use covert intervention as opposed to overt intervention in order to keep war limited.[167] Avoiding punishment or escalation may be the very reason that states choose covert influence operations instead of other overt tactics. Publicly attributing the operations to a foreign adversary may create pressure from domestic hawks to retaliate, increasing the potential for unwanted escalation. This applies to influence operations; operations waged covertly by foreign adversaries, if acknowledged by a target, may lead the target's population to pressure governmental decision-makers to retaliate.

Fear of escalation may be particularly acute because informational efforts may be perceived as relatively benign behavior on the spectrum of competitive state actions. There is a natural firebreak between informational efforts and the use of kinetic force. Decision-makers may view informational influence as a smaller threat than kinetic force, and thus fear escalating to a new, less desirable, category of activities should they choose to retaliate significantly.

---

[165] Kahn, Herman. 1965. *On Escalation: Metaphors and Scenarios*. London: Pall Mall Press.
[166] Shimer 2020.
[167] Carson 2018.

Figure 3: Why Defending Against Influence Operations is Challenging



Contributions of Influence Operations to International Security Studies

The three rationales presented above are policy-relevant reasons to study influence operations—effectively, to shape public discourse and to help understand the world as it is (and how it is evolving). Influence operations are prevalent and are likely to remain so because they are low in cost, because propagandists are likely to inflate their impact, and because they are difficult to deter. However, the study of influence operations can also contribute to the International Security Studies academic literature, expanding notions of state power and challenging traditional models of foreign policy decision-making.

Expanding Notions of State Power and Sharpening the Concept of Sharp Power

As a concept, power undergirds much of the International Relations literature. For Waltz, power differentiates between states that are functionally similar. It is a central concept in the literatures on balancing and bandwagoning,[168] on the balance of power,[169] and on offensive

[168] Walt, Stephen M. 1987. *The Origins of Alliances*. Cornell Studies in Security Affairs. Ithaca, New York: Cornell University Press.
[169] Paul, T.V., James J. Wirtz, and Michel Fortmann. 2004. *Balance of Power: Theory and Practice in the 21st Century*. Stanford, California: Stanford University Press; Wohlforth, William C., Richard Little, Stuart J. Kaufman, David Kang, Charles A. Jones, Victoria Tin-Bor Hui, Arthur Eckstein, Daniel Deudney, and William L. Brenner. 2007. "Testing Balance-of-Power Theory in World History." *European Journal of International Relations* 13(2): 155-185.

versus defensive realism.[170] Studying influence operations shows that the concept of power requires further development for 21st century statecraft.

The traditional notion of power in International Relations conforms to what Joseph Nye has described as "hard power." It is the ability of a country to get what it wants through coercion or payment and depends on a country's military and economic prowess. This conforms to a realist conception of power and is often operationalized in bean-count measures—with indicators such as population size, GDP, geography, and natural resources. "The logic of this approach is simple and sound: countries with more wealth and more military assets at their disposal tend to get their way more often than countries with fewer resources."[171] Beckley pushes back—contending that net indicators better measure state power than gross ones, since countries with lower liabilities but the same amount of resources have greater capacity to influence events or fight than do higher liability states.[172]

Public opinion plays a limited role in the existing account of hard power. Much of this work treats states as rational unitary actors, which makes variation in public opinion functionally unimportant. Hard power literature that does not rely on the rationality of a unitary state could consider public opinion in-depth: whether a state can be coerced or whether concessions can be purchased depend on the preferences of citizens. While some work on economic sanctions and attrition warfare considers this explicitly, the role of public opinion is treated as relatively unimportant in discussions of hard power.[173]

By contrast, public opinion is central to a second conception of power articled by Joseph Nye in 1990. In his introduction of "soft power," Nye made a theoretical contribution by distinguishing the logic of coercion and payment from the logic of attraction and persuasion. Dahl wrote that "My intuitive idea of power, then, is something like this: *A* has power over *B* to

---

[170] Mearsheimer, John. 1994. "The False Promise of International Institutions." *International Security* 19(3): 5-49. versus Glaser, Charles L. 1997. "The Security Dilemma Revisited." *World Politics* 50(1): 171-201; Jervis, Robert. 1978. "Cooperation under the Security Dilemma." *World Politics* 30(2): 167-214; Van Evera 1990.

[171] Beckley, Michael. 2018. "The Power of Nations: Measuring What Matters." *International Security* 43(2): 8. See also: Kennedy, Paul. 1987. *The Rise and Fall of the Great Powers.* Penguin.

[172] Beckley 2018.

[173] For examples on sanctions, see: Allen, Susan Hannah. 2008. "The Domestic Political Costs of Economic Sanctions." *The Journal of Conflict Resolution* 52(6): 916–944; Frye, Timothy. 2019. "Economic Sanctions and Public Opinion: Survey Experiments From Russia." *Comparative Political Studies* 52(7): 967-994; Grossman, Guy, Devorah Manekin, and Yotam Margalit. 2018. "How Sanctions Affect Public Opinion in Target Countries: Experimental Evidence From Israel." *Comparative Political Studies* 51(14): 1823–1857. For an example on public opinion in attrition warfare, see: Holden-Reid, Brian. 2011. "The American Civil War and The Strategy Of Attrition, 1861–65." *The RUSI Journal*. 156(3): 88-95.

the extent that he can get *B* to do something that *B* would not otherwise do."[174] Hard power, or coercion, is one way for A to get B to do something that B would not otherwise do. So, too, is soft power, or the power of attraction. Nye wrote that soft power "arises from the attractiveness of a country's culture, political ideals, and policies. When our policies are seen as legitimate in the eyes of others, our soft power is enhanced."[175] Others have refined this conception and argued that "*public views about current foreign policy*—rather than underlying public affinity—[sic] are directly relevant in shaping international outcomes."[176] Whether we consider a broader conception of soft power based on cultures, values, institutions and past policies, or a narrower one based on public views about current foreign policy, public opinion is central to the concept of soft power. Persuasion and attraction are dependent on the opinions of the other.

While the introduction of soft power is a welcome contribution in showing how information consumed by foreign publics is a form of power, influence operations show that Nye's dichotomy does not reflect the range of tools of statecraft.[177] Consider a social media influence operation, in which State A attempts to interfere in the domestic politics of State B by promoting a particular candidate or policy. If citizens of State B take actions they otherwise would not, this reflects a form of power according to Dahl's definition. However, it reflects neither the hard power of coercion (since it is unknown to B), nor the soft power of attraction. Studying influence operations captures a category of power that is neither overt persuasion nor overt coercion—it takes a covert form and can have disparate aims.

The closest articulation to this concept is Ludwig and Walker's notion of "sharp power," introduced in 2017. Sharp power describes activities by authoritarian states that fit neither the hard nor soft power categories. Sharp power, Ludwig and Walker write, "pierces, penetrates, or perforates the political and information environments in the targeted countries."[178] The authors contend because of the domestic governance of authoritarian leaders and their goals in foreign

---

[174] Dahl, Robert A. 1957. "The Concept of Power." *Behavioral Science* 2(3): 202-203.
[175] Nye, Joseph S. 2004. "Soft Power and American Foreign Policy." *Political Science Quarterly* 119(2): 255-270.
[176] Goldsmith and Horiuchi 2012.
[177] Morgenthau recognized a spectrum of tools by which one can exert power over another: "Power may comprise anything that establishes and maintains control of man over man. Thus, power covers all social relationship which serve that end, from physical violence to the must subtle psychological ties by which one mind controls another" (Morgenthau 1978). Lukes suggests that power is not merely winning a conflict or setting the agenda, but it can also be influencing the preferences of others. If influence operations allow states to shape the interests and behaviours of others, then influence operations can augment or compose a state's power.
[178] Walker, Christopher and Jessica Ludwig. 2017. "Sharp Power: Rising Authoritarian Influence" *National Endowment for Democracy Report.*

affairs, their activities cannot be seen as soft. Studying influence operations can sharpen the "sharp power" debate, and help to build analytical concepts of power that rely on public opinion but do not take the form of persuasion.

The sharp power literature requires further development and clarification. The authors seem to apply the term as a catch-all for a large range of activities that seem undesirable and to move between defining sharp power based on the perpetrator, tactic, and motivation.

First, if sharp power is meant to be a form of power—a way that A gets B to do what B otherwise would not—defining the form of power based on the nature of the actor is problematic. Historically, both democratic and authoritarian states have pierced the political and information environments of target states. It is unclear why an identical activity by a democracy would be soft power, but by an autocracy it would be sharp power.

Second, defining sharp power based on the tactic would be reasonable if the definition was clear and consistent. However, Ludwig and Walker include state-media and people-to-people exchange programs as forms of sharp power when used by Russia and China, not state-media and people-to-people exchange programs when engaged in covert manipulation. Thus, they revert to equating action with actor. An alternative could be to distinguish between soft power and sharp power based on an overt-covert dichotomy, as Nye does when he weighs in on sharp power.[179] By this definition, public diplomacy or white propaganda would qualify as soft power, while influence operations would qualify as sharp.

Third, defining a type of power based on specific motivations is challenging. Consider that soft power of attraction and hard power of coercion can be used for a range of aims; specifying that sharp power refers to activities designed to distract or to undermine independent institutions, for example, leaves open how to consider covert influence for other aims. States use influence operations to promote their own image, which may not be hostile to the target.

The evolution of literature has shown the beginning of a move to recognize that State A can get State B to do what State B would otherwise would not through means beyond military power. Soft power, or attraction, is relevant. And, as influence operations show, so is a category that is neither hard nor soft. Scholars could build on the few studies on sharp power to create a stronger theorization for forms of power that are neither hard nor soft—as the capacity to manipulate foreign discourse or covertly influence elite behavior is a form of power.

[179] Nye, Joseph S. 2019. "Protecting Democracy in an Era of Cyber Information War." *Belfer Center Paper*.

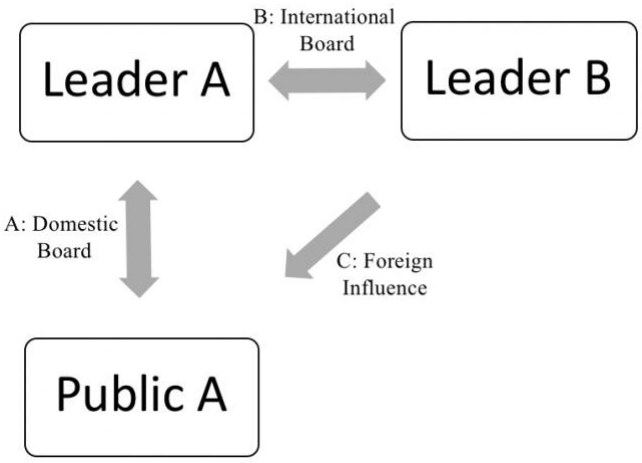
Challenging Foreign Policy Decision-Making Models

A second way that influence operations can contribute to International Security Studies is by highlighting a missing dimension in canonical accounts of foreign policy decision-making models. Standard models in International Relations—such as Allison's three models and Putnam's two-level game—conceive of domestic and international interactions as distinct. For Allison, Model 1 takes place on an international board where states are rational black boxes. Model 2 and Model 3, by contrast, take place on the domestic level, where intra-national organizations and leaders jockey for their favored positions.[180] Putnam articulates a two-level game: at the national level, domestic groups vie for their favored policies and politicians gain power by creating coalitions of groups; on the international level, governments "seek to maximize their own ability to satisfy domestic pressures, while minimizing the adverse consequences of foreign developments."[181] In both cases, the position of public opinion, domestic groups, and the leaders of departments and agencies is treated as *ex ante* given.

Influence operations challenge the notion that domestic and international boards are as disconnected as they appear in these models; foreign leaders, through influence operations, can influence the domestic board—independent of the domestic leader. Both Allison's models and Putnam's two-level games do not provide room for foreign governments to talk to, or influence, domestic audiences—independent of the state's leader. In the stylized diagram below, Allison and Putnam consider Pathway A and Pathway B, but do not take into account the potential for Pathway C.

[180] Allison 1971.
[181] Putnam, Robert D. 1988. "Diplomacy and Domestic Politics: The Logic of Two-Level Games." *International Organization* 42(3): 434.

Figure 4: Stylized Model of Domestic and International Interactions



Threat Perception, Influence Operations, and Tenets of Security Studies

The missing dimension from common models—Pathway C in the diagram above—has significant implications for theories of international security. For example, dominant conceptions of threat perception in international politics—both in rationalist and psychological accounts—focus on the state or leader level of analysis (on Pathway B). In the rationalist camp, Fearon situates his work on rationalist explanations of war from the perspective of a leader. His central question is: What prevents *state leaders* from reaching a negotiated settlement that would avoid the costs of war?[182] Threat perception is based on perception of capabilities and future intentions, and war may be rational from a leader's perspective because of (1) private information, (2) commitment problems, and (3) issue indivisibility.[183] In Fearon's model, state leaders are rational and thus functionally interchangeable. In the psychological strand, Jervis focuses on reasons why leaders misperceive the intentions of others.[184] While Jervis highlights common tendencies (including based on heuristics and biases), the psychological literature allows for greater variation between leaders[185]

A newer literature in International Relations suggests that public opinion matters to leaders (Pathway B in the diagram above). For example, audience cost literature suggests that

[182] Emphasis mine.
[183] Fearon, James D. 1995. "Rationalist explanations for war." *International Organization* 49(3): 379-414.
[184] Jervis 1976. Jervis, Robert. 2017. *How Statesmen Think: The Psychology of International Politics*, *How Statesmen Think*. Princeton University Press.
[185] Jervis 1976; Gross Stein, Janice. 1988. "Building Politics into Psychology: The Misperception of Threat." *Political Psychology* 9(2): 245-271.

leaders face a price for making threats and backing down.[186] Using public opinion surveys, Tomz finds evidence for audience costs across a range of conditions.[187] More recently, Tomz, Weeks, and Yarhi-Milo found that members of the Israeli parliament were "more willing to use force when the public was in favor and believed that contravening public opinion would entail heavy politics costs."[188] If this is the case, then public opinion can also shape leaders' threat perceptions and willingness to use force.

Studying influence operations demonstrates that these two literatures provide only partial accounts. Threat perception takes place at the leader level (Pathway B) and at the public level. The public can be influenced, or influence, domestic leaders (Pathway A), but it can also be addressed directly (and covertly) from foreign actors (Pathway C). Without Pathway C, existing accounts discount that public opinion cannot be treated as *ex ante* given—the model is dynamic and public opinion can be shaped by foreign leaders directly (and covertly).

This impacts multiple concepts of International Security. For example, threat perception undergirds theories of arms races and alliance formation. If State A perceives a threat from State B, State A is likely to build up its arms and/or invest in alliances. Studying influence operations and placing greater emphasis on Pathway C suggests that foreign actors can influence a competitor's threat perception via its domestic public—altering its proclivity for internal balancing. In regards to alliance formation, if foreign leaders can influence what domestic audiences think about their intentions and capabilities, they can influence the extent to which publics are willing to make concessions necessary for alliances. Or, foreign leaders can attempt to influence domestic publics' perceptions of third-party states, using influence operations as a wedge strategy between potential allies.[189] The feasibility of influence operations shaping perceptions of threat may be conditioned on offense-defense distinguishability: when offensive and defensive weapons or steps are highly distinguishable and when intelligence collection is easy/confident, influence operations are unlikely to have a significant effect compared to objective assessments. However, when weapons are non-distinguishable and motivations are unclear, influence operations may have a bigger effect.

---

[186] Fearon 1994.
[187] Tomz, Michael. 2007. "Domestic Audience Costs in International Relations: An Experimental Approach." *International Organization* 61(4): 821-840.
[188] Tomz, Weeks, and Yarhi-Milo 2020, 119.
[189] Wigell, Mikael. 2019. "Hybrid Interference as a Wedge Strategy: A Theory of External Interference in Liberal Democracy." *International Affairs* 95(2): 255–275.

Forging Ahead: A Research Agenda

In this paper, I have argued that influence operations—or "covert informational efforts intended to affect attitudes or opinions in a manner favorable to the perpetrator"—are underexplored within International Security Studies. How to define the subject remains fraught with controversy, and they are not well studied in top journals of Political Science, International Relations, and International Security. Nevertheless, they merit attention for both their policy significance and their ability to contribute to the Security Studies discipline. In this section, I conclude by outlining several potential research avenues.

First, scholars could look at variation in state usage. Why do some states invest more heavily in waging influence operations than others? Why do some states wage different types of influence operations than others? Researchers might start with a typology of influence operations—such as Lin and Kerr's distinction between propaganda operations, leak operations, and chaos-producing operations—and then analyze variation in usage of these tactics by state.[190] Potential explanations could include regime type, a state's norms, domestic public support (or opposition) to waging operations, and a state's power position vis-à-vis strategic dyads. This would help policymakers better assess which types of states are likely to wage influence operations (and of what nature) in order to defend against them.

Second, scholars should continue to explore the effectiveness of influence operations. What types of operations are more effective than others?[191] Under what conditions are influence operations likely to be effective? Scholars working on these questions can build on work in psychology on belief in misinformation and work in Political Science on cue-taking—applying existing lessons to this domain. Further research might examine the effects of over- and under-estimating the effectiveness of influence operations. Over-estimating influence operations might foster paranoia and mistrust, allowing the perpetrator to sow discord even if the underlying influence operation did not achieve its objective. It also can lead those in target countries to endorse the use of influence operations. Research suggests that belief in the effectiveness of terrorism increases support for its use.[192] At the same time, under-estimating the effect of

[190] Lin and Kerr 2019.
[191] Cormac et al. 2021.
[192] Houck, Shannon C., James McFarland, Laura V. Machia, and Lucian Gideon Conway III. 2019. "When Beliefs Lead to (Im)Moral Action: How Believing in Torture's Effectiveness Shapes the Endorsement of Its Use." *Political Psychology* 40(6): 1315–1339.

influence operations may lead policymakers to ignore them or not take the steps necessary to defend their state against foreign interference.

Third, research has begun to assess the trajectory of influence operations over time. There are many claims about how the introduction of the Internet, the state of globalization, and the tools at states' disposals allow them to wage influence operations differently than in the past. At the same time, others note that little is unprecedented when it comes to foreign interference.[193] Research can build on historical cases of influence operations to assess how modern influence operations are similar and different. Recent studies of social media and International Relations provide a strong starting point.[194]

[193] Kelton, Mark. "Why 'Unprecedented' Is Never the Right Word for Russian Election-Meddling," June 18, 2017, *The Cipher Brief*. Accessed: https://www.thecipherbrief.com/column_article/why-unprecedented-is-never-the-right-word-for-russian-election-meddling

[194] E.g. Kreps, Sarah. 2020. *Social Media and International Relations*. Cambridge University Press.

Appendix

Prevalence of Studies of Influence Operations in Top Journals

I conduct a systematic study of whether influence operations have been studied within top journals in the International Security Studies field. Methodologically, I follow a number of "studies of studies" of the discipline by analyzing subject mentions over time in top journals.[195] I present support for an empirical fact: that influence operations are not well-studied in top journals in Political Science, International Relations, or International Security Studies.

Most relevant to the subject of this article, Amy Zegart argues that U.S. intelligence is understudied in academia. She finds that between 2001 and 2006, the top three journals (*American Political Science Review, American Journal of Political Science,* and *Journal of Politics*) in Political Science published 750 articles and only one examined intelligence. Within International Relations journals, the top four journals (*International Organization, World Politics, International Studies Quarterly,* and *International Security)* included only five articles that covered the subject.[196] Intelligence is relevant to the question at hand because, as Rid notes in his recent work on active measures, "active measures are not spontaneous lies by politicians, but the methodical output of large bureaucracies. Disinformation was, and in many ways continues to be, the domain of intelligence agencies—professionally run, continually improved, and usually employed against foreign adversaries."[197] From Zegart's work, we should assume that influence operations have not been a focus of the discipline. However, to draw stronger conclusions, an independent search with more specific terms is necessary.

I build on Zegart's model by distinguishing between Political Science-wide journals and top journals within International Relations. I also include a third category: Security Studies sub-field journals. For each journal, I searched for articles that include five key terms in the text: influence operation(s), information operation(s), psychological operation(s), psychological warfare, and political warfare. For the first three terms, I conducted independent searches for the

[195] For an example of this type of study, Munck and Snyder assess the direction of comparative politics by constructing a dataset of articles from three leading U.S. journals (*Comparative Political Studies*, *Comparative Politics*, and *World Politics*) from 1989 to 2004 and coding for factors such as subject and theory generation versus theory testing. See: Munck, Gerardo L. and Richard Snyder. 2007. "Debating the Direction of Comparative Politics: An Analysis of Leading Journals." *Comparative Political Studies* 40(1): 5-31.
[196] Zegart 2007, p. 22.
[197] Rid 2020, p. 9.

singular and plural. From search results, I excluded book reviews, front matter, and back matter, as well as articles that include the term in a footnote or reference list but not the body text.

While *influence operation* is not a new term, I broaden my search to a number of related terms because diverse nomenclature is often used to describe the same event or tactic. At the same time, articles that include terms that overlap—but are not interchangeable—with influence operations may capture a different phenomenon. For example, political warfare may be used to describe an influence operation, but it is also often used to describe domestic political maneuvers that are unrelated to international security. Likewise, although some scholars use influence operations and information operations interchangeably,[198] information operations within military circles are discrete from influence operations. Joint Publication 3-13 "Information Operations," includes that the Secretary of Defense now characterizes information operations as "the integrated employment, during military operations, of IRCs in concert with other lines of operation to influence, disrupt, corrupt, or usurp the decision making of adversaries and potential adversaries while protecting our own" (2014, ix). Some of these activities—such as defending our own information capabilities—do not fall within influence operations as defined in this study. Influence operations can also take place outside of military operations. Thus, while this search strategy may provide additional insight, the inclusion of other terms (i.e. information operation/s, political warfare, psychological warfare) in the article does not necessarily imply a focus on influence operations.

Since articles mention terms that are tangential to their focus, I consider all articles that mention a term once as well as articles that mention a term several times. The search strategy described above produces 184 article-terms. Influence operation is the least popular term, returning only four articles; psychological warfare is the most popular, returning 78 articles. Note that the unit of analysis is an article that mentions a term, and so a single article that mentions multiple terms will be counted repeatedly. This coding strategy biases towards over-counting articles—creating a harder case for my argument that influence operations are under-studied. However, when we consider the number of articles that mention a search term multiple times, the number drops drastically: 31 articles mention a term at least twice, though only one of these articles uses influence operation; 16 articles use a search term at least three times and

[198] E.g. Gomez, Miguel Alberto 2020. "Cyber-Enabled Information Warfare and Influence Operations." *Information Warfare in the Age of Cyber Conflict*. Routledge.

influence operation is not among terms mentioned at least three times in any article. The low number of studies shows that even if influence operations were studied under a different term, they are underexplored in the International Relations literature.

Table 4 shows the years for each journal included in the search, as well as the database through which I executed the search. Due to limitations with journals' websites, the search does not include all issues up to 2020 for two journals: it includes *American Political Science Review* articles through 2014, and *International Organization* articles through 2016.

Table 4: Journals and Year Range Included in Search

Journal Categorization	Journal Title	JStor	Journal Website	Years Included in Search
Top Political Science Journals	American Journal of Political Science	1973-2018	AJPS Website: 2000-2020	1973-2020
	American Political Science Review	1906-2014	APSR website does not return bigram search for body text	1906-2014
	Journal of Politics	1939-2015	1939-2020	1939-2020
Top International Relations Journals	International Security	1976-2014	IS Website: 1996-2020	1976-2020
	International Organization	1947-2016	IO website does not return bigram search for body text	1947-2016
	International Studies Quarterly	1967-2014	1967-2020	1967-2020
Top Security Studies Journals	Journal of Conflict Resolution	1957-2016	JCR Website: 1957-2020	1957-2020
	Security Studies	Not on JStor	SS Website: 1991-2020	1991-2020

Table 5: Number of Articles That Mention Key Terms At Least Twice, By Journal

		Influence Operation(s)	Information Operation(s)	Political Warfare	Psychological Operation(s)	Psychological Warfare
Political Science Journals	American Journal of Political Science	0	0	0	0	0
	American Political Science Review	0	0	2	0	2
	Journal of Politics	0	0	0	0	1
Top Journals in International Relations	International Organization	0	0	1	0	1
	International Studies Quarterly	0	0	1	2	1
	International Security	0	3	0	1	2
Top Journals in Security Studies	Security Studies	1	5	0	5	1
	Journal of Conflict Resolution	0	0	0	0	2

Table Comparing Influence Operations to Related Key Terms

I use the four criteria that define political influence operations to compare the topic to related concepts. Since there is not a clear body of literature on influence operations within Political Science or International Relations, it is important to show that the category of influence operations is not identical to another category of statecraft. I show that related concepts are under- or over-inclusive, justifying the use of influence operations as a distinct category. By demonstrating this along three criteria, I also provide a replicable framework to distinguish between other related concepts.

In Table 6, I compare influence operations to propaganda, espionage, open-source intelligence (OSINT), psychological operations, covert action, and hybrid warfare based on the four criteria used to define influence operations in the paper's definitional section. For each term, I specify whether the category has or does not have each of the three requirements to classify as an interstate influence operation: informational, designed to affect attitudes or opinions, covert.

If a column is marked with a + or -, it indicates that, by definition, the concept has (or does not have) the criterion in question. For example, espionage is a form of intelligence collection, and, by definition, is not designed to affect opinion. It therefore is marked with a – under "Designed to Affect Opinion," and would not classify as an influence operation. By contrast, propaganda is marked with a +/- in both concealment and foreign; propaganda can be overt, and it can be domestic. So, some types of propaganda will qualify, while others—those that are domestic or those that are overt—will not classify as an influence operation. This table is not meant to be exhaustive, but to illustrate differences between influence operations and several related subjects.

Table 6: Influence Operations and Related Terminology

Related Topic	Informational	To Affect Opinions/ Attitudes	Covert	Foreign	Example from Category that Doesn't Fit Influence Operations
Influence Operations	+	+	+	+ / -	
Propaganda	+	+	+ / -	+ / -	The United States accompanied its Marshall Plan aid in Italy with films intended to portray the United States as a model others could obtain through the democratic way of life (Whelan 2010)
Espionage	+	-	+	+ / -	Christopher John Boyce and Andrew Daulton Lee sold Soviets sensitive data related to U.S. satellite systems (The Unseen Conflict—Foreign Espionage Operations Against the United States)
Open-Source Intelligence Collection	+	-	+ / -	+ / -	Bellingcat used open-source intelligence methods to study the Boogaloo movement
Psychological Operations	+ / -	+	+ / -	+	The United States dispersed leaflets in Korea "reminding the enemy of the superiority of the United Nations military forces, the futility of resistance, and the preferable option of surrender." (Jacobson diss, 6.)
Covert Action	+ / -	+ / -	+	+	Germany, Italy, and the Soviet Union covertly intervened in the 1936-1939 Spanish Civil War without public acknowledgement (Carson 2018, 13)
Hybrid Warfare	+ / -	+ / -	+ / -	+ / -	Russia encouraged demonstrations in Ukraine, used "little green men" to occupy government buildings, and oversaw a referendum to give legitimacy to annexation (Lanoszka 2016, 175)

Chapter 3: Follow the Experts or Follow the Crowd? Experimental Evidence on Public Attitudes Towards a Cyber-Attack

Dissertation Paper 2

Josh A. Goldstein[1]

Abstract

Top-down models of public opinion formation in International Relations posit that citizens have limited information about foreign policy and thus form their foreign policy preferences by following their favored elites. These models pose two puzzles for the discipline. First, actors waging influence operations often attempt to appear as non-elite members of the target state when trying to influence mass publics. Why would governments attempt to influence the foreign policy preferences of citizens in other countries through peer and public pressures (e.g. by forging public opinion polls or posing as locals) if non-elite influences play a marginal role in shaping opinions? Second, why does public opinion about foreign policy issues sometimes diverge from elite opinion, even when elites are united? Building on recent bottom-up advances, this paper suggests that despite the conventional wisdom, non-elite bottom-up cues can play a role in shaping foreign policy preferences. Leveraging data from two cross-country surveys in the United States and United Kingdom, I test the relative effect of public and expert cues on respondents' support for the use of force in response to a hypothetical cyber-attack. Expert and public cues both affect respondent preferences, and expert cue effects do not significantly outweigh public ones. In addition to supporting recent bottom-up advances in a cross-country context, this paper highlights the need to explore the mechanisms through which different cues operate and suggests that influence operations that work through falsified peer and public cues may impact foreign policy preferences.

[1] The project received ethical approval from the University of Oxford's IRB under Research Ethics Approval No. SSH_DPIR_C1A_19_039. Funding for survey experiments was provided by St. Hilda's College, Oxford, Oxford's Department of Politics and International Relations, Nuffield College, and the Smith Richardson Foundation. I thank Valerie Bradley, Jemma Conner, Janina Dill, Andrew Eggers, Michael Gross, Serena Hagerty, Dominic Johnson, Joel Kariel, Joshua Kertzer, Spyros Kosmidis, Soledad Artiz Prillaman, and Andrea Ruggeri for advice and feedback at various stages of the project. I also thank participants at the Oxford Political Psychology Research Cluster, the Changing Character of War Working Group, and the International Studies Association 2021 Annual Conference.

The "Bandwagon Effect" and the Puzzle of the Elite-Public Divide

Dominant models of public opinion formation in International Relations (IR) posit that members of the public have limited knowledge of foreign affairs, and thus form their preferences by following elite cues. When elites are united, public opinion is expected to follow elite views, whereas when elites diverge, public opinion is predicted to fall along partisan lines.[2] The conventional wisdom on public opinion formation helps to explain shifts in public support for U.S. intervention during World War II and Iraq.[3] However, it offers puzzles for both International Security Studies and International Relations.

First, for the International Security Studies community, if public opinion is elite-determined, why would governments attempt to influence the foreign policy preferences of foreign citizens through peer and public cues? For instance, in March 1940, A.A. Dudley of the British Library of Information sent a memo to the British Embassy in Washington, D.C. proposing that the British capitalize on the "bandwagon effect of public opinion polls" through the "use of these polls in a favourable direction, especially in helping Americans to remember that they are opposed to Nazism…"[4] By publicizing the results of leading polling questions, the British sought to sway American opinion away from the isolationist stance vis-à-vis the war on the European continent. Attempting to influence foreign citizens via peer and public pressures is a common form of influence operation. This strategy relies on an "operational effectiveness" logic, where a perpetrator believes their message is more likely to be effective if they conceal their identity.[5] Today, social media influence operations frequently leverage accounts that are designed to appear indigenous to the target state and interact with genuine social media users.[6] Either influence operations that rely on peer and public pressures to affect public opinion are

---

[2] Berinsky, Adam J. 2009. *In Time of War: Understanding American Public Opinion from World War II to Iraq*. Chicago Studies in American Politics. Chicago: University of Chicago Press; Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.

[3] Berinsky, Adam J. 2007. "Assuming the Costs of War: Events, Elites, and American Public Support for Military Conflict." *The Journal of Politics* 69(4): 975–97.

[4] The National Archives (TNA), "Letter to the Embassy in regard to public opinion polls in America," March 19th, 1940, FO 371/24228/265-268.

[5] See Dissertation Paper 1.

[6] Martin, Diego A., Jacob N. Shapiro, and Julia G. Ilhardt. 2020. "Trends in Online Influence Efforts." *Working Paper: Version 2.0*. August 5, 2020. Accessed: https://scholar.princeton.edu/sites/default/files/jns/files/trends_in_online_influence_efforts_v2.0_aug_5_2020.pdf See also: Goldstein, Josh A. and Lindsey Hundley. 2021. "Viva le #Resistance? Now-Suspended Twitter Accounts from Iran Claim #Resistance Personas in US Politics and Middle Eastern Affairs." *Stanford Internet Observatory Report*.

doomed to fail, or the top-down view of public opinion formation in International Relations discounts the role of non-elite cues.

Second, for International Relations scholars, if public opinion is elite-determined, what explains historical instances in which elites are in consensus but public opinion is divided? In 2014, why did the U.S. public divide in blaming Israel and Hamas for violence in Israel when the elites were in consensus in blaming Hamas?[7] Or, during the intervention in Afghanistan, why did citizens in countries contributing to the NATO-led mission, such as the United Kingdom, Germany, France, Italy and Canada, oppose the intervention when their elites supported it?[8] A potential answer is that the top-down model oversells the role of elite cues. Group dynamics, peer opinions, or even public opinion polls present an alternative source of bottom-up cues.[9] Instead of simply following elites, members of the public may also follow their peers or the public writ large. This helps to explain recurring elite-public divides and suggests that influence operations from foreign entities that operate via peer and public cues are not destined to fail.

Unfortunately, with few exceptions, the International Relations literature has not directly compared the effect of elite cues and non-elite peer or public cues on individuals' foreign policy preferences.[10] In this article, I rely on survey experiments in the United States and United Kingdom to isolate and directly compare how knowledge of the purported view of (a) national security experts and (b) the public affects respondents' foreign policy preferences. If the top-down model assumed in much of the literature holds, the effect of elite cues should be much larger than the effect of non-elite cues (e.g. public opinion polls). If, however, public cues are relatively sizable, it challenges the top-down model, provides further support to Kertzer and

---

[7] Kertzer, Joshua D. and Thomas Zeitzoff. 2017. "A Bottom-Up Theory of Public Opinion about Foreign Policy." *American Journal of Political Science* 61(3): 543–58.

[8] Kreps writes that the top-down view of public opinion formation is "remarkably at odds with the experience of states participating in Afghanistan operations…where elite consensus has done little to improve the negative public environment." Kreps, Sarah. 2010. "Elite Consensus as a Determinant of Alliance Cohesion: Why Public Opinion Hardly Matters for NATO-Led Operations in Afghanistan." *Foreign Policy Analysis* 6(3): 191–215.
For other examples of elite-public divides, see: Busby, Joshua W. and Jonathan Monten. 2012. "Republican Elites and Foreign Policy Attitudes." *Political Science Quarterly* 127(1): 105–42; Kay, Sean, and Patrick Cronin. "Bridging America's Foreign Policy Elite-Main Street Divide." *War on the Rocks*, August 11, 2016. Accessed: https://warontherocks.com/2016/08/bridging-americas-foreign-policy-elite-main-street-divide/; Kriner, Douglas, and Graham Wilson. 2016. "The Elasticity of Reality and British Support for the War in Afghanistan." *The British Journal of Politics and International Relations* 18(3): 559–80; Busby, Joshua, Craig Kafura, Jonathan Monten, and Jordan Tama. 2019. "Multilateralism and the Use of Force: Experimental Evidence on the Views of Foreign Policy Elites." *Foreign Policy Analysis* 16(1): 118-129.

[9] Crutchfield, Richard S. 1955. "Conformity and Character." *American Psychologist* 10(5): 191-198: 191–98; Cialdini, Robert B. 2009. *Influence: Science and Practice*. 5th ed. Boston: Pearson/Allyn and Bacon.

[10] Kertzer and Zeitzoff 2017.

Zeitzoff's argument regarding the significance of bottom-up mechanisms and citizens' underlying beliefs, and foregrounds a new set of questions regarding public opinion about foreign policy. It also suggests that the operational effectiveness logic of secrecy in influence operations may be worth pursuing—as peer influences can produce attitudinal change.

To test these competing logics, I ran two cross-national survey experiments in the United States and United Kingdom in which I presented respondents with a vignette of a cyber-attack on the electric grid of a major city in their home country, varied the cue treatments, and asked respondents about their support for a retaliatory strike. The domain of the attack is notable for both methodological and policy reasons. From a methodological standpoint, the domain provides a "most likely" case for expert cues given asymmetric information between the public and national security experts.[11] The cyber domain should also allow for strong cue effects given the lack of partisan and experiential priors (i.e. most Americans and Brits lack direct experience with cyber-attacks on their electric grids, and such attacks have not, to-date, become politicized). From a policy perspective, public support for retaliation is particularly likely to be policy-relevant in this domain given that cyber "lacks clear rules of engagement or elite consensus on what constitutes an 'act of war.'"[12] Yet, the experimental literature on public attitudes on the use of force in response to cyber-attacks is nascent.[13] Thus, in addition to contributing to the top-down versus bottom-up debate, this article provides data on public support for kinetic retaliations to cyber-attacks.

The results from the survey experiments challenge the top-down conventional wisdom. First, I study the effects of expert and public cues as independent instruments. When experts or

---

[11] Eckstein, Harry. "Case Study and Theory in Political Science." 2009. In *Case Study Method*, edited by Roger Gomm, Martyn Hammersley, and Peter Foster. United Kingdom: SAGE Publications Ltd; George, Alexander L., and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. BCSIA Studies in International Security. Cambridge: MIT Press.

[12] Kreps, Sarah, and Jacquelyn Schneider. 2019. "Escalation Firebreaks in the Cyber, Conventional, and Nuclear Domains: Moving beyond Effects-Based Logics." *Journal of Cybersecurity* 5(1): 2.

[13] See: Kreps, Sarah, and Debak Das. 2017. "Warring from the Virtual to the Real: Assessing the Public's Threshold for War over Cyber Security." *Research & Politics* 4(2); Jensen, Benjamin, and Brandon Valeriano. 2019. "What Do We Know about Cyber Escalation? Observations from Simulations and Surveys." *Atlantic Council: Scowcroft Center for Strategy and Security;* Tomz, Michael and Jessica L. P. Weeks. 2020. "Public Opinion and Foreign Electoral Intervention." *American Political Science Review* 114(3): 856-873; Dvir, Rotem, Nehemia Geva, and Arnold Vedlitz. 2021. "Unpacking Public Perceptions of Terrorism: Does Type of Attack Matter?" *Studies in Conflict & Terrorism*: 1-21; Shandler, Ryan, Michael L. Gross, Sophia Backhaus, and Daphna Canetti. 2021. "Cyber Terrorism and Public Support for Retaliation: A Multi-Country Survey Experiment." *British Journal of Political Science*: 1-19; Shandler, Ryan, Michael L. Gross, and Daphna Canetti. 2021. "A Fragile Public Preference for Cyber Strikes: Evidence from Survey Experiments in the United States, United Kingdom, and Israel." *Contemporary Security Policy* 42(2): 135-162.

the public oppose using force in response to a cyber-attack, respondents are less likely to support a retaliatory strike. Contrary to expectations based on the conventional literature, the size of the two effects is comparable.[14] Second, building on work by Kertzer and Zeitzoff,[15] I study the effect of presenting both expert and public cues to respondents. When the cues diverge from one another (i.e. experts support the use of force but the public don't, or the public support the use of force but experts don't), expert cues do not significantly outweigh public ones. Instead, the cues offset one another and support for retaliation does not significantly differ from a control group with no information on expert or public cues. When the cues are in consensus (i.e. both experts and the public oppose a retaliatory strike), average support for retaliation is not significantly different than when only one of the two cues opposes a retaliatory strike. Hence, elite and non-elite cues have a substitutive effect, perhaps because they play similar roles in providing respondents with a second opinion of what constitutes a reasonable answer about the proposed retaliation.

This article proceeds in five steps: I begin by summarizing existing research on public cue-taking about foreign policy, paying particular attention to the top-down logic and its critics. Then, I present a competing, bottom-up model and outline hypotheses for the direction, magnitude, and interaction of public and expert cues. Third, I introduce the experimental design and highlight the significance of the domain of the vignette—a cyber-attack on the electric grid of a major city. Fourth, I present survey findings, before conducting exploratory analysis on the mechanisms through which the cues operate, using additional dependent variables beyond support for retaliation. Finally, I conclude with a discussion of future directions for research on public opinion formation about foreign policy and the different mechanisms of covertly influencing foreign publics.

Previous Research on Cue-Taking and Public Opinion about Foreign Policy

The top-down model of public opinion formation that dominates the International Relations literature evolved in response to the Almond-Lippmann consensus, which took a skeptical view of public opinion about foreign policy. The Almond-Lippmann consensus

[14] In the U.K. samples, the expert and public cues are both significant in both experiments. In the U.S. samples, which are relatively less well-powered than the U.K., the expert cue is significant in the first experiment but not the second, and the public cue is significant in the second experiment but not the first.
[15] Kertzer and Zeitzoff 2017.

suggested that members of the public do not pay attention to foreign policy issues, that public opinion is volatile in response to new information, and that the mass public lack clear ideologies to structure their beliefs.[16] Research on the Vietnam War challenged this view by showing that the American public held consistent views—they were reluctant to pay the costs of war, strongly supported negotiations with the Vietcong, and were responsive to American soldier casualties.[17] Stronger support for the consistent view position came in the 1980s and early 1990s when polling organizations traced public opinion on the same set of foreign policy issues over time and found a high level of consistency.[18]

A common explanation that reconciled the fact that people know little about foreign policy issues and yet still hold reasonable opinions is that people use elite heuristics as informational shortcuts. While earlier scholars had taken a skeptical view of informational shortcuts,[19] more recently researchers have argued that, through the use of heuristics, people can "be knowledgeable in their reasoning about political choices without necessarily possessing a large body of knowledge about politics."[20] Given that it may be rational for any one person to "minimize his investment in political information,"[21] Neuman describes this top-down process as a "natural and effective division of labor."[22]

---

[16] Almond, Gabriel A. 1950. *The American People and Foreign Policy*. 1st ed. New York: Harcourt; Converse, Philip E. 2006. "The nature of belief systems in mass publics" (1964), Critical Review. 18:1-3, 1-74. For an overview, see: Holsti, Ole R. 2004. *Public Opinion and American Foreign Policy*. Rev. Analytical Perspectives on Politics. Ann Arbor: University of Michigan Press.

[17] Verba, Sidney, Richard A. Brody, Edwin B. Parker, Norman H. Nie, Nelson W. Polsby, Paul Ekman, and Gordon S. Black. 1967. "Public Opinion and the Vietnam War" *The American Political Science Review* 61(2): 317-333; Mueller, John E. 1971. "Trends in Popular Support for the Wars in Korea and Vietnam." *American Political Science Review* 65(2): 358-375; Mueller, John E. 1973. *War, Presidents, and Public Opinion*. New York: Wiley; Gartner, Scott Sigmund. 2008. "The Multiple Effects of Casualties on Public Support for War: An Experimental Approach." *American Political Science Review* 102(1): 95–106.

[18] Shapiro, Robert Y. and Benjamin I. Page. 1988. "Foreign Policy and the Rational Public." *Journal of Conflict Resolution* 32(2): 211–47; Page, Benjamin I. and Robert Y. Shapiro. 1992. *The Rational Public: Fifty Years of Trends in Americans' Policy Preferences*. American Politics and Political Economy. Chicago: University of Chicago Press. For a strong overview of this literature, see: Eichenberg, Richard C. 2016. "Public Opinion on Foreign Policy Issues." In *Oxford Research Encyclopedia of Politics*, by Richard C. Eichenberg. Oxford University Press.

[19] Lippmann, Walter. 1955. *Essays in the Public Philosophy*. New York: New American Library; McGuire, William J. 1969. "The Nature of Attitudes and Attitude Change." In *The Handbook of Social Psychology*, edited by Gardner Lindzey and Elliot Aronson, 2nd ed. Reading, MA: Addison-Wesley.

[20] Sniderman, Paul M., Richard A. Brody, and Philip E. Tetlock. 1991. *Reasoning and Choice: Explorations in Political Psychology*. Cambridge: Cambridge University Press. p. 19.

[21] Downs, Anthony. 1957. "An Economic Theory of Political Action in a Democracy." *Journal of Political Economy* 65(2): 135–50.

[22] Neuman, W. Russell. 1986. *The Paradox of Mass Politics: Knowledge and Opinion in the American Electorate*. Cambridge, Mass: Harvard University Press. p. 186.

In addition to economy of attention, there are additional reasons why members of the public may take cue from elites. First, even if the public is well-informed, elites may form their views based on "inside information" to which the public would otherwise not have access.[23] Second, a robust literature in human evolutionary biology suggests that there is a deeper, evolutionary advantage to selective cultural learning. Humans have been more successful than other mammals at adapting to novel environments because of a unique ability to learn from others.[24] People do not randomly select from who they learn (i.e. they do not imitate), but instead are selective about who they turn to for learning based on who demonstrates success (skill bias), who has accumulated symbols of success (success bias), and whom others turn to for learning (prestige bias).[25]

Applied to the field of foreign policy, citizens, wired to learn selectively from others, may be particularly likely to consider the view of those with domain-relevant expertise. Elites are likely to benefit from each of the three biases: their language indicates knowledge (skill bias), their pedigree indicates success (success bias), and the notion that they are public—that they opine on governmental policy, or that they have been selected to serve in positions of influence—suggests that others have deferred to them too (prestige bias). Throughout this article, I follow Busby et al. in defining foreign policy elites as the "community of individuals with professional experience or expertise related to foreign policy and international affairs" and refer to foreign policy elites and experts interchangeably.[26]

These theoretical arguments are supported by a robust literature examining the effect of different elite cues. Zaller demonstrates that political elites play a major role in setting the terms

---

[23] Carmines, Edward G. and James H. Kuklinski. 1990. "Incentives, opportunities, and the logic of public opinion in American political representation." In J.A. Ferejohn and J.H. Kuklinski (eds.), *Information and Democratic Processes*, pp. 240–268. Urbana: University of Illinois Press.

[24] Boyd, Robert, Peter J. Richerson, and Joseph Henrich. 2011. "The Cultural Niche: Why Social Learning Is Essential for Human Adaptation." *Proceedings of the National Academy of Sciences of the United States of America* 108: 10918–10925.

[25] Boyd, Robert, and Peter J. Richerson. 1985. *Culture and the Evolutionary Process*. Chicago: University of Chicago Press; Chudek, Maciej, Sarah Heller, Susan Birch, and Joseph Henrich. 2012. "Prestige-biased cultural learning: Bystander's differential attention to potential models influences children's learning." *Evolution and Human Behavior* 33(1): 46-56.

[26] Busby et al 2019. Based on this definition, elites include both domain-specific policy experts and partisan elected officials. This differs from the narrow conception of elite presented by Levendusky and Horowitz. See: Levendusky, Matthew S. and Michael C. Horowitz. 2012. "When Backing Down Is the Right Decision: Partisanship, New Information, and Audience Costs." *The Journal of Politics* 74(2): 323-338.
I select a broader definition of elites because non-elected officials can have an influence on public opinion. In the experimental design, I test one type of elite (foreign policy experts), but further research could examine the relative effect of peer or public cues with different types of elite cues.

of public discourse and framing issues, which has a strong effect on mass opinion.[27] "Faced with a confusingly large number of issues," he writes, "many citizens offer opinions based on the positions of the political figures they most trust."[28] Studies have found that public opinion has followed presidential policy on issues including "the liberation of Kuwait, the invasion of Grenada, the testing of nuclear weapons, relations with the People's Republic of China, isolationism, and both the escalation and de-escalation of the Vietnam War."[29]

Both Zaller and Berinsky highlight that consensus support for war among U.S. government officials increases public support, whereas disunity among leaders leads citizens to follow partisan lines.[30] Other lines of research add nuances of different types to the elite-driven model, arguing that the effect of a cue depends on whether the source is favorable, trustworthy and credible;[31] whether the issue is familiar or unfamiliar, scientific or emotional;[32] whether the cue receiver has high or low levels of political knowledge;[33] and how issues are framed.[34]

Based on the theoretical and empirical support for top-down elite cues, I predict:

> *Hypothesis 1: Follow the Experts: Positive (negative) endorsement of a foreign policy course of action by national security experts will have a positive (negative) effect on a respondent's support for the policy.*

---

[27] Zaller 1992.
[28] Zaller, John. 2012. "What Nature and Origin Leaves Out." *Critical Review* 24(4): 569–642.
[29] Edwards, George C. 2003. *On Deaf Ears: The Limits of the Bully Pulpit*. New Haven: London: Yale University Press. p. 28.
[30] Zaller 1992; Berinsky 2009; Larson, Eric V. and Bogdan Savych. 2015. *American Public Support for U.S. Military Operations from Mogadishu to Baghdad*. Santa Monica, CA: RAND.
[31] Hovland, Carl I. and Walter Weiss. 1951. "The Influence of Source Credibility on Communication Effectiveness." *The Public Opinion Quarterly* 15(4): 635–650; McGuire 1969; Carmines and Kuklinski 1990; Mondak, Jeffery J. 1993. "Source Cues and Policy Approval: The Cognitive Dynamics of Public Support for the Reagan Agenda." *American Journal of Political Science* 37(1): 186-212; Druckman, James N. 2001. "On the Limits of Framing Effects: Who Can Frame?" *The Journal of Politics* 63(4): 1041-1066; Maoz, Ifat, Andrew Ward, Michael Katz, and Lee Ross. 2002. "Reactive devaluation of an "Israeli" vs."Palestinian" peace proposal." *Journal of Conflict Resolution* 46(4): 515-546.
[32] Carmines, Edward G. and James A. Stimson. 1980. "The Two Faces of Issue Voting." *American Political Science Review* 74(1): 78–91; Guisinger, Alexandra, and Elizabeth N. Saunders. 2017. "Mapping the Boundaries of Elite Cues: How Elites Shape Mass Opinion across International Issues." *International Studies Quarterly* 61(2): 425–41.
[33] Zaller 1992; Kam, Cindy. 2005. "Who Toes the Party Line? Cues, Values, and Individual Differences." *Political Behavior* 27(2): 163–82.
[34] Entman, Robert M. 2004. *Projections of Power: Framing News, Public Opinion, and U.S. Foreign Policy*. Studies in Communication, Media, and Public Opinion. Chicago: University of Chicago Press; Kam 2005; Baum, Matthew A. and Tim Groeling. 2010. "Reality Asserts Itself: Public Opinion on Iraq and the Elasticity of Reality." *International Organization* 64(3): 443–79.

An Alternative to the Top-Down Logic: Peer and Public Effects

The top-down model of opinion formation can be critiqued on at least two fronts: first, elite cues may not provide the "natural and effective division of labor" that Neumann described,[35] and second, the empirical record does not always conform to the theory's expectations.

First, there are reasons to be skeptical of the claim that elite cues provide a cure to the lack of foreign policy knowledge among the general public. A body of research has challenged the position that elite cues allow members of the public to hold preferences "as-if" they were informed. If elites seek to deceive the public, cue-taking comes at a cost of democratic accountability and could lead people to make choices contrary to their true preferences.[36] Or, if partisan elites hold positions that are incongruent with partisan stereotypes, citizens may fail to adjust their candidate preferences if they do not question intra-party elite cues.[37]

A second line of critique challenges the empirical evidence claiming to show that public opinion follows elite opinion. In a comprehensive study on "going public," Edwards finds that "presidents typically do not succeed in their efforts to change public opinion."[38] In a study of elite and public attitudes between 1974 and 2002, Page and Jacobs find at least a 10-percentage point difference between elite and public preferences on 73% of the 1,153 survey questions asked.[39] As in the case of participant countries in Afghanistan and 2014 public opinion in the United States over violence in Israel described in the introduction of this paper, the elite-public gap suggests that the public does not simply take cues from the top.

---

[35] Neumann 1986.

[36] Gibbs, David N. 1995. "Secrecy and International Relations." *Journal of Peace Research* 32(2): 213–28; Mearsheimer, John J. 2011. *Why Leaders Lie: The Truth about Lying in International Politics*. New York, N.Y.: Oxford University Press; Sagar, Rahul. 2013. *Secrets and Leaks: The Dilemma of State Secrecy*. Princeton, NJ: Princeton University Press; Schuessler, John M. 2015. *Deceit on the Road to War: Presidents, Politics and American Democracy*. Cornell Studies in Security Affairs. Ithaca: Cornell University Press.

[37] Kuklinski, James H., and Norman L. Hurley. 1994. "On Hearing and Interpreting Political Messages: A Cautionary Tale of Citizen Cue-Taking." *The Journal of Politics* 56(3): 729–51; Gabel, Matthew, and Kenneth Scheve. 2007. "Estimating the Effect of Elite Communications on Public Opinion Using Instrumental Variables." *American Journal of Political Science* 51(4): 1013–28; Arceneaux, Kevin. 2008. "Can Partisan Cues Diminish Democratic Accountability?" *Political Behavior* 30(2): 139–60; Lenz, Gabriel S. 2009. "Learning and Opinion Change, Not Priming: Reconsidering the Priming Hypothesis." *American Journal of Political Science* 53(4): 821–37.

[38] Edwards 2003, 241.

[39] Page, Benjamin I., and Lawrence R. Jacobs. 2006. "A Disconnect Between Policy Makers and the Public?" In *The Foreign Policy Disconnect: What Americans Want from Our Leaders but Don't Get*, by Benjamin I. Page and Marshall M. Bouton. Chicago, Ill.: University of Chicago Press.

While theories including prestige bias and selective learning account for why humans may take cues from experts, there are equally plausible theories for why individuals may take cues from non-elite sources. Social psychology research identifies multiple pathways through which knowledge of group opinion influences the individual towards conformity, including: (1) normative influence processes such as the natural human desire to belong ("to be wanted, liked, accepted, and approved by others"), and (2) informational influence processes, which "revolve around the equally human drive to know, to be correct, and to understand how best to act in a certain situation."[40]

The normative influence theory is supported by psychology research. Tajfel and Turner famously argue that social categorizations "provide a system or orientation for *self*-reference: they create and define the individual's place in society." Since individuals identify with broader social groups and seek to maintain high self-esteem, they seek to hold positive images of those groups as well.[41] Thus, if equipped with knowledge of group opinion, they may bring their own opinion into line with that of their group to maintain a positive identity.[42] By contrast, the informational influence process may be reflected in what Cialdini terms the "Principle of Social Proof"—the notion that "we determine what is correct by finding out what other people think is correct."[43] Group opinion signals to a receiver what others think is correct, allowing an individual to benefit from wisdom of the crowds.[44]

As in the case of elite cues, there is convincing evidence across domains that people are influenced by peer and public cues. Asch famously tested the effect that majority opinion has on an individual's opinion when the majority opinion is incorrect. When a group unanimously insists that the length of one stick matches another even though it clearly does not, a substantial

[40] McDermott, Rose. 2004. *Political Psychology in International Relations*. Analytical Perspectives on Politics. Ann Arbor: The University of Michigan Press. p. 84. See also: Deutsch, Morton, and Harold B. Gerard. 1955. "A Study of Normative and Informational Social Influences upon Individual Judgment." *The Journal of Abnormal and Social Psychology* 51(3): 629–36; Rothschild, David, and Neil Malhotra. 2014. "Are Public Opinion Polls Self-Fulfilling Prophecies?" *Research & Politics* 1(2): 1-10.
[41] Tajfel, Henry, and John Turner. 1979. "Chapter 3 An Integrative Theory of Intergroup Conflict." In *The Social Psychology of Intergroup Relations*, edited by William G. Austin and Stephen Worchel. p. 40.
[42] Cialdini, Robert B. and Noah J. Goldstein. 2004. "Social Influence: Compliance and Conformity." *Annual Review of Psychology* 55: 591-621.
[43] Cialdini 2009.
[44] Surowiecki, James. 2004. *The Wisdom of Crowds: Why the Many Are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. London: Little Brown.

minority of test subjects yield their judgment in accordance with the majority.[45] In a study of foreign policy preferences, Gorden recorded private opinions about Russia held by individuals in a housing cooperation in the United States. When individuals then shared their opinion in a group setting, they tended to "conform to their conception of the group norms."[46] Likewise, Rothschild and Malhotra find that "polls, by directly influencing individual-level support for policies, can be self-fulfilling prophecies and produce opinion cascades."[47] Based on these theoretical and empirical studies, I predict:

> *Hypothesis 2: Follow the Group: Positive (negative) endorsement of a foreign policy course of action by public opinion will have a positive (negative) effect on a respondent's support for the policy.*

The Experts versus The Crowd

Thus far, we have two theories which generate hypotheses that both expert and group opinions affect public foreign policy preferences. Each theory is grounded in work from multiple academic disciplines and has support from existing historical and experimental studies. Yet, these literatures largely stand apart. To date, this author has identified only one experiment in the field of public opinion about foreign policy that has probed how these two cues interact and weighed them comparatively within a single experiment. Kertzer and Zeitzoff's "A Bottom Up Theory of Public Opinion about Foreign Policy" makes a significant contribution to the literature by developing a bottom-up theory to challenge the dominant paradigm. It also provides room for further study.[48]

Kertzer and Zeitzoff present American respondents with mock news clippings about policymakers in Washington, D.C. debating a salient national security issue, such as a potential military pivot in response to a rising China or troop deployment to the Middle East.[49] In each

---

[45] Asch, Solomon E. 1951. "Effects of Group Pressure upon the Modification and Distortion of Judgments." In *Groups, Leadership and Men; Research in Human Relations*, edited by Harold Steere Guetzkow, 177–90. Pittsburgh, PA: Carnegie Press.
[46] Gorden, Raymond L. 1952. "Interaction between Attitude and the Definition of the Situation in the Expression of Opinion." *American Sociological Review* 17(1): 58.
[47] Rothschild and Malhotra 2014, p. 6. See also: Ansolabehere, Stephen and Shanto Iyengar. 1994. "Of Horseshoes and Horse Races: Experimental Studies of the Impact of Poll Results on Electoral Behavior." *Political Communication* 11(4): 413–30; Marsh, Catherine. 1985. "Back on the Bandwagon: The Effect of Opinion Polls on Public Opinion." *British Journal of Political Science* 15(1): 51–74.
[48] Kertzer and Zeitzoff 2017.
[49] Ibid p. 547.

case, an elite cue about the proposed policy (operationalized as support from Congress members or a Congress member of a particular party) was embedded within the vignette. After reading the news story, respondents were shown a graph with the purported percentage of other survey respondents who supported the policy under consideration. Finally, respondents were asked about their own level of support. The paper finds that group influence was at least as large as expert influence across issue sets, which makes a strong argument that the top-down model in International Relations requires revisiting.[50]

Yet, generalizations regarding top-down versus bottom-up influences based on these experiments alone are limited in two ways. First, the experimental design may pose a "hard case" for the top-down cue. Kertzer and Zeitzoff operationalize the elite cue as whether a Democrat or Republican in Congress (or Democrats or Republicans in Congress) support a proposed action and imbed the cue within the vignette. The bottom-up cue is operationalized as the (purported) support or opposition of other survey respondents, and comes after the respondent reads the news clip.[51] Their peer cue is a second-order judgment, as it indicates how other people judged the scenario (which included the elite cue). This is likely to create a strong bottom-up cue, particularly compared to an elite cue that is embedded within the news clip itself.

The second limitation is that because only one study has directly compared bottom-up and top-down cues, our understanding of relative effect is limited to the cues they selected. We cannot generalize to other types of bottom-up cues (such as public opinion polls) and other types of top-down cues (such as the view of the foreign policy establishment) based on experimental manipulations of the purported view of other survey respondents and of Congress.

In the survey to follow, I build on Kertzer and Zeitzoff's novel findings by offering an easier case for the elites (the most established theory) and a harder case for the public in cross-country experiments. Thus, if the public cue is relatively strong even under these conditions, it strengthens the Kertzer and Zeitzoff model and further implies that the top-down models that dominate the International Relations literature need rethinking. I create this harder test for the

[50] Kertzer and Zeitzoff 2017.

[51] The use of Congress as the top-down cue may create a hard case for the top-down model as well, since most members of Congress are not national security experts (for an account by a member of Congress, see Pascrell Jr. 2019) and Congress members may not be perceived by the public favorably. During the dates in which the survey was deployed, approval ratings of Congress did not pass 20% and disapproval ratings did not drop below 74%. However, this is at least partially mitigated by the fact that the authors describe the Congress person as a "veteran member of Congress with established foreign policy expertise" and include party affiliation.

public in several ways: First, instead of using Congress for the elite cue, I use "national security experts" to ensure domain-specific expertise is conveyed and the potential stigma of Congress removed. Second, instead of using a second-order judgment for a peer effect, I provide a harder case for the bottom-up cue by using a public opinion cue imbedded within the vignette parallel to the elite cue to maximize comparability. I tell respondents that according to a recent poll the majority of national security experts support (do not support) an action, and the majority of the American/British public support (do not support) an action. In the real world, people may be faced with scenarios closer to the Kertzer and Zeitzoff design: given opinions of peers who have knowledge of Congress persons' preferences. Thus, I sacrifice a level of external validity for internal validity in putting the bottom-up and top-down cues on equal footing.

If Kertzer and Zeitzoff (2017) provide a relatively hard case for elites, we cannot yet be certain regarding the relative importance of either variable or the relative validity of the competing models presented above. This is especially true given that there are good theoretical reasons to expect either public *or* elite cues to be the stronger of the two.

The public may win out given the lack of contemporary belief in, or reverence for, elites. According to Nichols, the zeitgeist of the age includes a "Google-fueled, Wikipedia-based, blog-sodden collapse of any division between professionals and laymen, students and teachers."[52] This trajectory can be traced to factors including the rise of social media and mass-publishing and the failures of the elite class to provide sound policy advice about issues such as the Iraq War or the 2008 financial crisis. Not surprisingly, the Edelman Trust Barometer recorded all-time low levels of global trust in 2017.[53] Moreover, skeptics of elite decision-making may be justified: in certain domains, experts fail to outperform novices. From 1988 to 2003, Tetlock asked 284 experts to forecast outcomes across political, economic, and national security issues. The study finds that "collapsing across all judgments, experts on their home turf made neither better calibrated nor more discriminating forecasts than did dilettante trespassers."[54] Moreover, Tetlock concludes that "experts who assigned confidence estimates of 80 percent or higher were correct only 45

[52] Nichols 2017.
[53] Edelman. "2019 Edelman Trust Barometer: Global Report." Accessed: https://www.edelman.com/trust/2019-trust-barometer
[54] Tetlock, Philip E. 2005. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, N.J.: Princeton University Press. p. 54. See also: Bradley, James V. 1981. "Overconfidence in Ignorant Experts." *Bulletin of the Psychonomic Society* 17(2): 82–84.

percent of the time."[55] The combination of expert incompetence and overconfidence may lead people to reject expert cues in favor of the crowd.

By contrast, people may be more influenced by elites than the public—whether or not it is justified. First, individuals may heed the advice of experts out of the belief that elites have an informational advantage. Members of the general public (i.e. non-foreign policy experts) know they lack information on relevant details. Humans tend to have a "theory of mind," or the ability to explain and predict other people's behavior by attributing independent mental states (such as beliefs and desires) to them.[56] As a result of the information gap, citizens may be more likely to trust experts, who are perceived as having domain-specific, decision-relevant knowledge.

Second, individuals may trust experts more than the public *even if* experts have no better ability to make foreign policy predictions than the average citizen. In other words, the public may attribute to the experts expertise that they do not possess. A study of stock market predictions found that stock market professionals and laypeople made similar sized errors in predictions. Yet, both groups predicted that the errors by laypeople would be approximately double the size of those by the experts.[57] In some domains, people have greater trust in experts than themselves even when the belief is unwarranted. In the experimental test to follow, the domain of the vignette—a cyber-attack on the electric grid of a major city—may be perceived as technical and accessible to experts, creating a likely domain for expert cues.

Due to the evidence for the top-down and bottom-up models, I posit competing hypotheses for effect strength of the two cues:

> *Hypothesis 3a: The Bottom-Up Hypothesis: Public opinion will have a stronger effect than expert opinion if the two diverge.*
>
> *Hypothesis 3b: The Top-Down Hypothesis: Expert opinion will have a stronger effect than public opinion if the two diverge.*

[55] Tetlock 2005.
[56] Gallagher, Helen L. and Christopher D. Frith. 2003. "Functional Imaging of 'Theory of Mind.'" *Trends in Cognitive Sciences* 7(2): 77–83.
[57] Torngren, Gustaf, and Henry Montgomery. 2004. "Worse Than Chance? Performance and Confidence Among Professionals and Laypeople in the Stock Market." *Journal of Behavioral Finance* 5(3): 148–53.

Elite and Public Cues: Additive or Substitutive?

What happens when both the elites and the public agree? We could reasonably expect public and expert cues to have additive or substitutive effects. For example, if individuals form strong assessments based on the facts of the scenario, they may look to public and expert views simply to confirm their original assessment. In such a case, if either the experts or the public hold a belief that is congruent with their own, it would have a similar effect as if *both* the experts and the public confirmed their predisposition. On the other hand, the effect of two cues may be stronger than one because the intensity of the cue is stronger when the two groups agree or because the respondent may use the two cues in different ways (e.g. normative versus informational influences).

I thus propose competing hypotheses:

Hypothesis 4a: The Additive Hypothesis: Expert and public cues in consensus will have a stronger effect than either an expert or a public cue on its own.

Hypothesis 4b: The Substitutive Hypothesis: Expert and public cues in consensus will not have a significantly stronger effect than either an expert or a public cue on its own.

Summary of Hypotheses

Figure 1: Summary of Hypotheses for Elite and Public Cue Effects

Goal of Hypothesis	**Hypothesis Title**	**Hypothesis Description**
Direction of Cue Effects	Hypothesis 1: Follow the Experts	Positive (negative) endorsement of a foreign policy course of action by national security experts will have a positive (negative) effect on a respondent's support for the policy.
	Hypothesis 2: Follow the Crowd	Positive (negative) endorsement of a foreign policy course of action by public opinion will have a positive (negative) effect on a respondent's support for the policy.
Comparative Magnitude of Cue Effects	Hypothesis 3a: The Bottom-Up Hypothesis	Public opinion will have a stronger effect than expert opinion if the two diverge.
	Hypothesis 3b: The Top-Down Hypothesis	Expert opinion will have a stronger effect than public opinion if the two diverge.

Interaction of Cue Effects	Hypothesis 4a: The Additive Hypothesis	Expert and public cues in consensus will have a stronger effect than either an expert or a public cue on its own.
	Hypothesis 4b: The Substitutive Hypothesis	Expert and public cues in consensus will not have a significantly stronger effect than either an expert or a public cue on its own.

Experimental Design

In order to test the effect of expert and public cues on respondent foreign policy preferences, I deployed a series of survey experiments that manipulated expert and public support for a proposed action.

In Experiment 1, deployed with YouGov in July 2020 in both the United States and United Kingdom, respondents were asked to imagine a scenario in which there is a cyber-attack on the electric grid of a major American (British) city and the U.S. (British) government is debating whether to engage in an airstrike against the group responsible.[58] While the control group received this scenario, four treatment groups were told that according to a recent poll (a) the majority of national security experts support (do not support) the strike, and (b) the majority of the American (British) public supports (does not support) the strike. Experiment 1 allows us to test the relative effect of the two cues embedded within a single vignette, including when the two cues diverge (i.e. Treatment II and III in Figure 2).

Figure 2: Summary of Experiment 1: Control and Four Treatments

<table>
<tr><th rowspan="2">Control Group</th><td rowspan="3">+</td><td colspan="2" rowspan="2">(Public, Expert)</td><th colspan="2">Expert Cue</th></tr>
<tr><th>Expert Support</th><th>Expert Oppose</th></tr>
<tr><td rowspan="2">Receives Vignette (No Information on Expert or Public Opinion)</td><td rowspan="2">Public Cue</td><td>Public Support</td><td>Treatment I: Consensus Support: + / +</td><td>Treatment II: Disagreement: + / -</td></tr>
<tr><td>Public Oppose</td><td>Treatment III: Disagreement: - / +</td><td>Treatment IV: Consensus Opposition: - / -</td></tr>
</table>

[58] For full vignettes and experimental treatments, please see the Appendix.

In Experiment 2, fielded through YouGov in May 2019 and December 2019, U.S. and U.K. samples were presented with a similar underlying vignette to Experiment 1. While the control groups received no cues, three treatment groups were told that (a) the majority of the public do not support the strike, (b) the majority of national security experts do not support the strike, or (c) the majority of national security experts and the majority of the public do not support the strike. This allows for a clean test to examine how cues differ when presented individually and if the cue effects are additive or substitutive. It does not, however, include conditions of divergence.

Figure 3: Summary of Experiment 2: Control and Three Treatments

(Public, Expert)		**Expert Cue**	
		No Information	*Expert Oppose*
Public Cue	*No Information*	Control Group	Treatment II: Expert Opposition
	Public Oppose	Treatment I: Public Opposition	Treatment III: Consensus Opposition

After presenting respondents with the vignette, I asked respondents about the extent to which they support the government ordering an airstrike against the group responsible. I also asked three exploratory questions to investigate potential mechanisms of influence: whether the respondent believes an airstrike in response to the cyber-attack would likely deter future cyber-attacks, whether they believe an airstrike against the group responsible would be ethical, and how comfortable they would be sharing their response. (See Figure 4 for a summary, and the Appendix for the full questionnaire.) I present results of the exploratory analysis following the main results.

Importance of the Domain of Attack

I choose to test the relative influence of expert and public opinion with a fictional cyber-attack on an electric grid of a major city in particular for three reasons.

First, from a methodological standpoint, cyber-attacks on critical infrastructure compose a domain that has not been highly politicized and in which people lack strong priors. This has two effects: it makes the likelihood of cue-taking particularly strong, and it creates an "easy

case" for experts. Expert cues are likely to be stronger when polarization is low, when individuals lack prior attitudes on a topic, and when information is only presented once.[59] In this study, cue effects are likely to be particularly strong for experts (as compared to for the public) because (a) there may be a perceived information gap between experts and the public in the cyber domain, and (b) I explicitly operationalize the elite cue as "national security experts." If public effects are strong even in this experiment, it shows that even in a "hard case" Kertzer and Zeitzoff's bottom-up theory holds weight.

Second, public opinion formation is particularly important in instances involving a new use of military force, since it is a domain where American leaders may be attuned to domestic costs and pressures. For instance, fear of electoral repercussions "inhibited President Bush from launching a large-scale U.S. intervention [in Somalia] prior to November 1992."[60] This scenario speaks to a realistic unknown, as U.S. national security experts are increasingly discussing vulnerabilities to the U.S. electric grid. In 2018, Eric Rosenbach, former "cyber tsar" of the Defense Department, said that the United States is not ready for a major cyber-attack despite evidence that "much of the power grid has already been penetrated by Russian intelligence."[61] More recent reporting suggests that Russian hackers have "breached the power grid, water treatment facilities, and even nuclear power plants…" in the United States.[62] Zegart argues that the "cyber threats of tomorrow could disable the cars that we drive, the airplanes that we fly, they could turn off power or water to cities across the country for days or weeks or longer…"[63] The results of this survey experiment provide insight into the extent to which people will support a kinetic response to a cyber-attack and the potential for various cue givers to sway public opinion.

---

[59] Guisinger and Saunder 2017, p. 439; Kumkale, G. Tarcan, Dolores Albarracín, and Paul J. Seignourel. 2010. "The Effects of Source Credibility in the Presence or Absence of Prior Attitudes: Implications for the Design of Persuasive Communication Campaigns." *Journal of Applied Social Psychology* 40(6): 1325–1356.
[60] Baum, Matthew A. 2004. "How Public Opinion Constrains the Use of Force: The Case of Operation Restore Hope." *Presidential Studies Quarterly* 34(2): 187–226.
[61] Belfer Center, "U.S. Not Prepared for Major Cyber Attack." *NBC Today Show*, December 26, 2018. Accessed: https://www.belfercenter.org/publication/eric-rosenbach-nbc-us-not-prepared-major-cyber-attack.
[62] Perlroth, Nicole. "Russians Who Pose Election Threat Have Hacked Nuclear Plants and Power Grid." *The New York Times*. October 23, 2020. Accessed: https://www.nytimes.com/2020/10/23/us/politics/energetic-bear-russian-hackers.html
[63] As cited in: Ferguson, Niall. 2017. *The Square and the Tower: Networks and Power from the Freemasons to Facebook*. London: Penguin Books. p. 44.

Third, experimental studies of public opinion on support for the use of force in response to cyber-attacks are limited, due to the relatively new nature of the phenomenon. In a notable exception, Kreps and Schneider find that public support for a retaliatory strike is dependent on the method of the attack—people are less supportive of using force in response to cyber-attacks than nuclear or conventional attacks "even if they perceive the magnitude of attacks across domains to be comparable."[64] This challenges an effects-based logic and suggests cyber-attacks may provide an off-ramp from traditional escalation ladders. Gross et al. examine the psychological effects of cyber-terrorism in Israel and find that cyber-terrorism fosters fear akin to kinetic terrorism.[65] My experiment builds on this growing field of public opinion about cyber-attacks by testing whether elite and public cues may affect people's views of retaliatory measures in this domain.

Results

Experiment 1

Table 1 shows the effect of expert opposition (versus support) and public opposition (versus support) as independent instruments. This is a 2x2 design that only includes the treatment groups (see Figure 4).[66] Dropping the control group allows us to consider expert and public treatments each as binary variables.[67]

In the U.K. sample, expert opposition and public opposition each have small substantive effects (β=.225 and β=.270 on a five-point scale) that are statistically significant. In the U.S. sample, which is relatively less powered, the expert cue is significant but the public cue, which is substantively comparable, falls short of traditional levels of statistical significance. These results provide strong support for Hypothesis 1, as respondent support is affected by expert cues in both treatments, and mixed support for Hypothesis 2, as respondent support is affected by public cues at a significant level in the U.K. sample but not at a significant level in the U.S. sample.

[64] Kreps and Schneider 2019, p. 1.
[65] Gross, Michael L., Daphna Canetti, and Dana R. Vashdi. 2016. "The Psychological Effects of Cyber Terrorism." *Bulletin of the Atomic Scientists* 72(5): 284–91.
[66] In Figure 4, I provide the summary for the U.S. sample. The U.K. summary is identical, except the attack occurred against a British city and the dependent variables ask about a strike by the British government.
[67] In the Appendix, I reproduce the results and exclude the interaction term. Results from the robustness check are consistent with the findings in the main text.

Figure 4: Experiment 1 Design (Without Control)

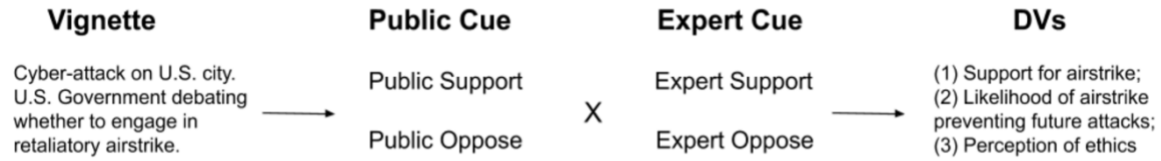


Table 1: Regression of Treatment on Support for Strike (Experiment 1)

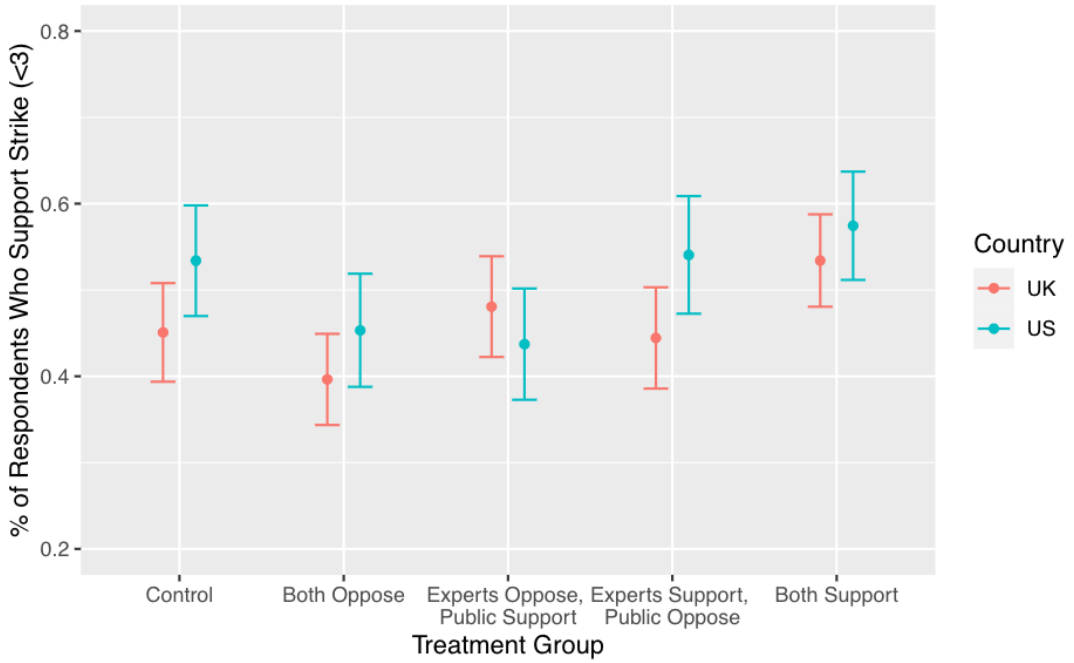
	Dependent variable:			
	Support for Strike			
	UK Sample (No Covs)	UK Sample (With Covs)	US Sample (No Covs)	US Sample (With Covs)
	(1)	(2)	(3)	(4)
Expert Oppose	-0.225^{**}	-0.218^{**}	-0.393^{***}	-0.375^{***}
	p = 0.041	p = 0.034	p = 0.003	p = 0.003
Public Oppose	-0.270^{**}	-0.250^{**}	-0.182	-0.195
	p = 0.015	p = 0.016	p = 0.178	p = 0.128
Age		0.009^{***}		0.014^{***}
		p = 0.0001		p = 0.00000
Gender		-0.356^{***}		-0.314^{***}
		p = 0.00001		p = 0.001
College		-0.053		-0.152
		p = 0.470		p = 0.115
Labour (UK)		-1.074^{***}		
		p = 0.000		
LibDem (UK)		-0.875^{***}		
		p = 0.000		
Other (UK)		-0.699^{***}		
		p = 0.000		
Republican (US)				1.151^{***}
				p = 0.000
Independent (US)				0.395^{***}
				p = 0.0005
Other (US)				0.179
				p = 0.486
ExpertOppose* PublicOppose	0.078	0.067	0.169	0.195
	p = 0.617	p = 0.644	p = 0.374	p = 0.278
Constant	3.323^{***}	3.806^{***}	3.574^{***}	2.647^{***}
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Observations	1,234	1,219	907	870
R^2	0.013	0.164	0.014	0.182
Adjusted R^2	0.010	0.157	0.011	0.174
Residual Std. Error	1.367 (df = 1230)	1.263 (df = 1209)	1.429 (df = 903)	1.315 (df = 860)
F Statistic	5.324^{***} (df = 3; 1230)	26.264^{***} (df = 9; 1209)	4.275^{***} (df = 3; 903)	21.320^{***} (df = 9; 860)
Note:	*p<0.1; **p<0.05; ***p<0.01			

Figure 5 shows the percent of respondents that supported a retaliatory strike by treatment group and country of sample. I coded supporters as those who responded above three on a five-point scale. In the U.K. sample, support is highest when both experts and the public support the strike (53.41% support the strike) and lowest when both the experts and the public oppose the

strike (39.63% support the strike). In the U.S. sample, support is highest when experts and the public support the strike (57.44% support the strike) and lowest when experts oppose the strike and public support it (43.72% support the strike). In this U.S. sample, however, the percent of supporters when experts oppose the strike and the public support it is close to the percent of supporters when both cues oppose the strike (45.33% support the retaliatory measure in the consensus opposition treatment). These results provide preliminary evidence for Hypotheses 1 and 2: When the public and experts are in agreement, respondent support follows the cues.

When the two cues diverge, results from Experiment 1 fail to adjudicate between Hypotheses 3a and 3b. Across both the U.S. and U.K. samples, average support in conditions when the two cues diverge is never significantly different from the control. When the two cues diverge, the effects offset one another, and one cue does not significantly outweigh the other. This defies the top-down conventional wisdom and suggests that bottom-up pressures can influence public opinion as well.

Figure 5: Percent of Respondents Who Support Strike by Treatment Group



Experiment 2

In Figure 7, I plot the percentage of respondents who supported the strike by treatment group for both U.S. and U.K. samples. In both the U.S. and U.K. samples, the percentage of

respondents who support the strike is highest in the control condition, in which respondents are given no information about expert or public opinion. When either the experts or the public oppose the strike, a smaller percentage of respondents support the strike than in the control. This provides support for Hypotheses 1 and 2; respondents appear to "Follow the Experts" and "Follow the Crowd" as fewer support the strike when the majority of (a) national security experts or (b) the public oppose it.

Table 2 presents regression models of cue treatments on respondent support for force in the U.S. and U.K. samples. In this case, both expert opposition and public opposition treatments have small and significant effects in the U.K. sample. In the U.S. sample, however, which is relatively less powered, the public effect is significant and the expert effect falls short of traditional levels of statistical significance. This challenges the top-down conventional literature, which would predict that expert cues would be relatively stronger than public ones.

Figure 6: Survey 2 Experimental Design

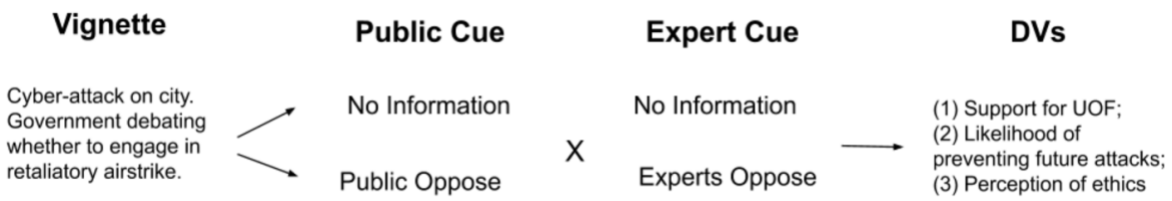


Figure 7: Percent of Respondents Who Support Strike by Treatment Group

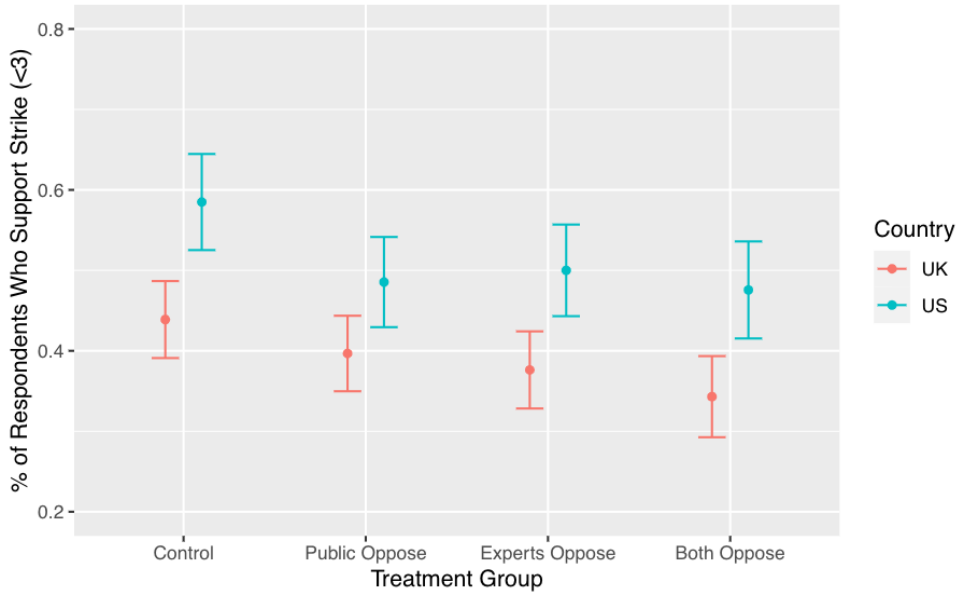


Table 2: Regression of Treatment on Support for Strike (Experiment 2)

	Dependent variable:			
	Support for Strike			
	UK Sample (No Covs)	UK Sample (With Covs)	US Sample (No Covs)	US Sample (With Covs)
	(1)	(2)	(3)	(4)
Public Oppose	−0.184**	−0.199**	−0.319***	−0.328***
	p = 0.046	p = 0.029	p = 0.006	p = 0.003
Experts Oppose	−0.232**	−0.318***	−0.148	−0.165
	p = 0.013	p = 0.001	p = 0.200	p = 0.126
Both Oppose	−0.300***	−0.352***	−0.266**	−0.227**
	p = 0.002	p = 0.0003	p = 0.026	p = 0.041
Age		0.007***		0.014***
		p = 0.0002		p = 0.000
Gender		0.470***		0.245***
		p = 0.000		p = 0.002
College		−0.277***		−0.211***
		p = 0.00005		p = 0.010
Labour (UK)		−0.766***		
		p = 0.000		
LibDem (UK)		−0.738***		
		p = 0.000		
Other (UK)				0.997***
				p = 0.000
Republican (US)				0.266***
				p = 0.006
Independent (US)				0.095
				p = 0.494
Other (US)		−0.444***		
		p = 0.00000		
Constant	3.106***	3.110***	3.558***	2.475***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Observations	1,578	1,452	1,141	1,141
R^2	0.007	0.137	0.008	0.156
Adjusted R^2	0.005	0.131	0.005	0.150
Residual Std. Error	1.328 (df = 1574)	1.250 (df = 1442)	1.372 (df = 1137)	1.269 (df = 1131)
F Statistic	3.678** (df = 3; 1574)	25.354*** (df = 9; 1442)	2.969** (df = 3; 1137)	23.288*** (df = 9; 1131)
Note:	*p<0.1; **p<0.05; ***p<0.01			

Finally, in order to test whether the two cues are additive or substitutive (i.e. Hypotheses 4a or 4b), I run four regression models in Table 3. The models show whether the effect of two cues in consensus (e.g. public and experts oppose) is statistically different from the effect of a single cue (e.g. public oppose only as referent, or experts oppose only as referent). Across countries, there is not a significant difference between treatments that are told only experts or only the public oppose a strike and treatments that are told *both* experts and the public oppose a strike. The effect of opposition by a single group reduces the effect of opposition by a second group, supporting Hypothesis 4b (The Substitutive Hypothesis).

Table 3: Regression of Treatment on Support for Strike (Experiment 2)

	Dependent variable:			
	Support for Strike			
	Public Oppose Referent (U.K.)	Expert Oppose Referent (U.K.)	Public Oppose Referent (U.S.)	Expert Oppose Referent (U.S.)
	(1)	(2)	(3)	(4)
Both Oppose	−0.150	−0.037	0.103	−0.089
	p = 0.120	p = 0.706	p = 0.339	p = 0.401
Age	0.007**	0.005*	0.014***	0.013***
	p = 0.026	p = 0.094	p = 0.00001	p = 0.00004
Gender	0.434***	0.414***	0.123	0.380***
	p = 0.00001	p = 0.00003	p = 0.258	p = 0.0004
College	−0.377***	−0.356***	−0.324***	−0.324***
	p = 0.0002	p = 0.0005	p = 0.006	p = 0.005
Labour (UK)	−0.865***	−0.778***		
	p = 0.000	p = 0.000		
LibDem (UK)	−0.857***	−0.681***		
	p = 0.00003	p = 0.0004		
Other (UK)	−0.494***	−0.419***		
	p = 0.0002	p = 0.002		
Republican (US)			1.102***	1.196***
			p = 0.000	p = 0.000
Independent (US)			0.423***	0.330**
			p = 0.003	p = 0.012
Other (US)			0.092	0.545***
			p = 0.616	p = 0.010
Constant	3.071***	2.979***	2.171***	2.259***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Observations	702	684	576	567
R^2	0.153	0.120	0.167	0.188
Adjusted R^2	0.144	0.111	0.157	0.178
Residual Std. Error	1.261 (df = 694)	1.276 (df = 676)	1.283 (df = 568)	1.248 (df = 559)
F Statistic	17.872*** (df = 7; 694)	13.152*** (df = 7; 676)	16.258*** (df = 7; 568)	18.518*** (df = 7; 559)
Note:	*p<0.1; **p<0.05; ***p<0.01			

Exploratory Analysis

Perceived Effectiveness and Ethics of Retaliation

The bulk of this article has focused on whether, and how much, expert and public cues affect support for the use of force. Based on the data presented above, both expert and public cues have small but significant effects. This spurs a second set of questions regarding the mechanism through which these two cues affect respondent support for the strike. The lacuna is particularly evident for public cues, which are not well theorized in International Relations.

To begin to probe how expert and public cues operate, I asked survey respondents two sets of questions. First, in addition to asking respondents about their support for a retaliatory strike, I asked respondents whether they would agree or disagree (on a five-point scale) with the following statements:

- An airstrike in response to the cyber-attack would likely deter future cyber-attacks.
- An airstrike against the group responsible for the cyber-attack would be ethical.

Returning to a distinction made by McDermott, social influence may occur through an information influence process, which revolves around the human drive to know, or a normative process, such as the natural human desire to belong.[68] If expert cues operate primarily through informational influence and social cues operate primary through normative influence, we might expect expert cues to have a larger effect on respondents' views about whether the airstrike would deter future cyber-attacks (a proxy for whether it would be effective) and public cues to have a larger effect on respondents' views about whether it would be ethical.

As I show in the Appendix, the results are mixed.[69] In Experiment 1, the expert opposition cue had an effect on perceived effectiveness of a retaliatory strike in the U.K. sample but not in the U.S. sample, and the public opposition cue did not have a significant effect on perceived effectiveness in either country. On the question of whether the strike would be ethical, neither the expert or public cue had a significant effect in either the U.S. or U.K samples. The weakness of these two cues may be due to the experimental design, as treatment respondents were always presented with two cues in Experiment 1.

In Experiment 2, where respondents were presented with single opposition cues or a consensus opposition treatment, public opposition significantly affected respondents' views of whether the strike would be ethical in both the U.S. and U.K samples. This is consistent with the idea that public cues operate through a normative influence pathway. On the question of deterrence, however, both cues affected respondents' views in the U.K. sample, and neither cue had a significant effect in the U.S. sample. Given the inconsistency of findings across country samples, future work could further unpack why expert and public support for the use of force affect respondents' views, and how expert and public cues differ based on country context.

[68] McDermott 2004.
[69] Full models for Experiment 1 samples are shown in Table 6 and Table 7 in the Appendix. Full models for Experiment 2 samples are shown in Table 17 and Table 18 in the Appendix.

Comfort Sharing Support for Retaliation

In addition to questions about perceived effectiveness and ethics, I developed a second test of normative influence. After the initial post-treatment questions (e.g. support, effectiveness, and ethics) I asked survey respondents:

> "How comfortable or uncomfortable would you be if the extent to which you supported the airstrike in the hypothetical scenario was reported on social media or in a local newspaper? (Your responses will remain anonymous. This is strictly hypothetical)."

If the public cue operates primarily through a normative pathway, we would expect respondents whose views diverge from public cues to be reluctant to share their responses. However, across treatment groups, those who supported the strike were consistently more willing to share their answers on a social media platform than the reference category (i.e. those who responded in the middle of the five-point scale). As shown in Figures 8 and 9, there was not a significant difference between those who opposed the strike and those who neither supported nor opposed the strike in most cases.

Figure 8: Experiment 1 Comfort Sharing Response by Treatment (U.S. and U.K. Samples)

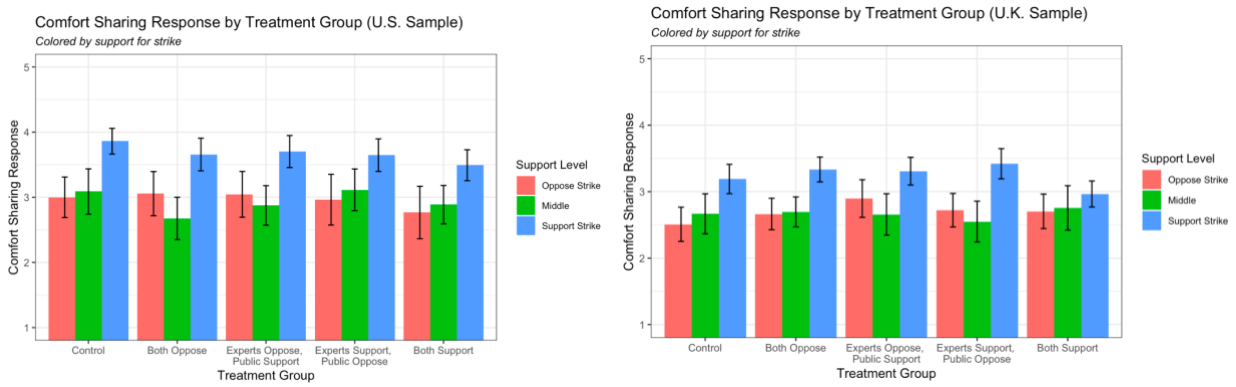
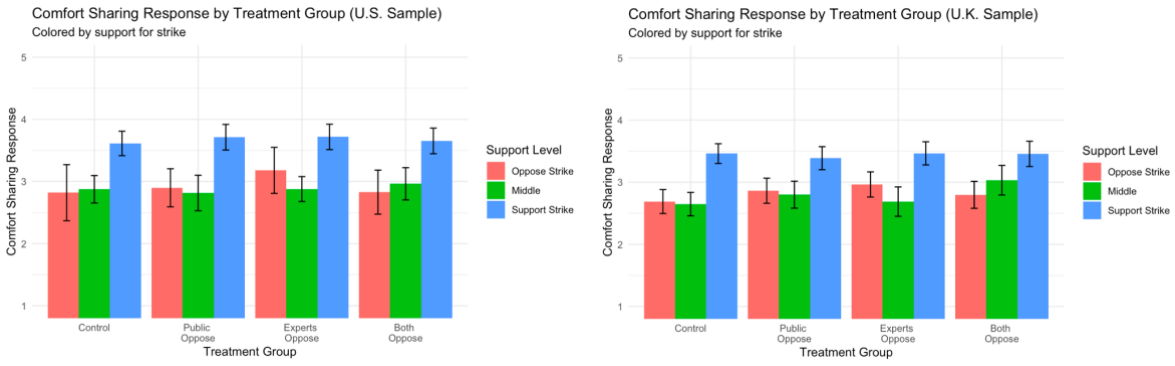


Figure 9: Experiment 2 Comfort Sharing Response by Treatment (U.S. and U.K. Samples)



The fact that respondents who support the strike are more comfortable sharing their response across all treatment groups in both countries in both experiments is notable. On the one hand, we would expect respondents who diverge from public cues to be less comfortable sharing their response in a public setting if cues operate in part through a normative pathway of a desire to fit in. This is particularly the case in the second set of experiments, where respondents in the first treatment group are presented with a single cue of public opposition. One possible explanation for the finding is that those who support the strike are more hawkish, and hawkishness may correlate with confidence.

Conclusion

Dominant models of public opinion formation in International Relations focus on the role of elite cues in shaping public opinion on foreign policy issues. With the exception of Kertzer and Zeitzoff's "bottom-up theory," the effect of bottom-up, non-elite cues has gone underexplored in this literature.[70] Yet, studies of conformity have long shown that people are influenced by peers and the public writ large when forming their preferences. Is the foreign policy domain an exception?

The results of two experiments involving both U.S. and U.K. samples suggest not. Respondents asked about their support for retaliation to a hypothetical cyber-attack take cues from both elites *and* the public. In the U.K. samples, public and expert effects were significant in both experiments. In the U.S. samples, which were relatively less well-powered, the expert effect was significant in the first experiment, and the public effect was significant in the second. Recall, as discussed above, that the cyber domain provides an easier case for national security experts given the information gap and the low level of political polarization over cyber-attacks on critical infrastructure. If the non-elite cue, operationalized as a soft "public opinion," played a significant role even in this context, it is likely to play at least as large a role in other ones that provide an easier case for the public. I find strong support for Kertzer and Zeitzoff's bottom-up model: bottom-up cues can impact people's foreign policy preferences as much as top-down elite ones.[71]

In addition to studying the effect of the two cues as independent instruments, I also tested the relative strength of the two cues when they diverge (in Experiment 1) and whether the effects

[70] Kertzer and Zeitzoff 2017.
[71] Ibid.

are additive or substitutive (in Experiment 2). In Experiment 1, I did not find that the expert cue or the public cue always outweighed the other. In both countries, in the two conditions in which the cues diverged, respondents' average support for retaliation did not differ significantly from the control group. This defies the logical extension of the conventional wisdom, which would suggest that a top-down cue of the beliefs of national security experts would likely outweigh a bottom-up public cue. In Experiment 2, I found that the effect of the two cues was substitutive. Respondents who were presented with opposition from a single cue (i.e. experts oppose or public oppose) did not have significantly different rates of support for retaliation compared to respondents who were told that both groups oppose the strike.

These findings have significant implications for the mechanisms of influence operations. As described in Paper 1, one of the logics for covert action in the informational domain is that the target will be more receptive of information when it comes from a deflective source. Foreign actors often pretend to be locals or share other fake peer cues (like public opinion polls) in their covert, informational campaigns. However, International Relations literature has largely ignored bottom-up influences. My findings suggest that the operational effectiveness logic of secrecy holds: If a foreign actor can successfully pose as peer or public cues, they may be effective in shaping respondents' foreign policy preferences.

Future studies could build on this work in a number of ways. First, researchers could experimentally vary different types of bottom-up cues. For instance, how does the effect of bottom-up cues differ when the bottom-up cue comes in the form of a public opinion poll, other people who read a news article, a social media post from a peer, or a citizen calling into a radio station? A recent review on the effects of influence operations wrote that "No study in this review directly tested for potential variations in effectiveness between different influence operations tactics."[72] Scholars might select bottom-up cues that closely align with different tactics of social media influence operations to help explain which types of operations are more or less likely to affect target attitudes. Second, further research could ask, under what conditions do public cues influence respondents' foreign policy preferences, and on which issues do public cues matter more or less than elite ones? Guisinger and Saunders theorize why elite messages

[72] Bateman, Jon, Elonnai Hickock, Laura Courchesne, Isra Thange, and Jacob N. Shapiro. 2021. "Measuring the Effects of Influence Operations: Key Findings and Gaps from Empirical Research." *Carnegie Endowment for International Peace*. Accessed: https://carnegieendowment.org/2021/06/28/measuring-effects-of-influence-operations-key-findings-and-gaps-from-empirical-research-pub-84824

move public opinion on some issues but not others.[73] The same could be explored for public cues, and for the relative effect of elite versus public ones. Findings would help the policy community understand the topics where the public is particularly susceptible to foreign influence operations, in order to determine where to prioritize limited resources. Third, scholars could also investigate the interaction between elite and non-elite cues and uncertainty. My models in this paper relied on perfect information, but the attribution of cyber-attacks is often uncertain. Further work could explore whether the effects of top-down versus bottom-up cues varies based on the attribution certainty. Finally, researchers could unpack the causal pathways through which peer and elite cues operate. This paper establishes that bottom-up cues about a particular foreign policy proposal impact citizens' support for that proposal. Future experiments could be designed to more critically determine why respondent support changes when presented with different cues.

[73] Guisinger and Saunders 2017.

Appendix: Questionnaires

Experiment 1 Vignette, Treatments, and Questions

Note: The vignette for American respondents in Experiment 1 was identical to the United Kingdom vignette except "United Kingdom" was replaced with "United States", and "British" was replaced with "U.S." or "American" depending on context.

Experiment 1 Vignette

Vignette: "The United Kingdom faces frequent cyber-attacks, with hostile countries responsible for many of the attacks.

Imagine that a cyberwarfare group in a hostile country enacts a cyber-attack against the electric grid of a British city. Residents are left without power for 3 days and 15 civilians die as their life-saving medical devices cannot be powered.

The British Government is debating whether to engage in an airstrike targeting the cyberwarfare group's headquarters in the foreign country, but the airstrike would probably result in at least 12 civilian casualties."

Treatments groups received a different concluding paragraph to the vignette:

- Control: [shown no information about expert or public opinion]
- Treatment 1: According to a recent poll, the majority of British national security experts and the majority of British citizens do not support a retaliatory strike.
- Treatment 2: According to a recent poll, the majority of British national security experts do not support a retaliatory strike but the majority of British citizens support the strike.
- Treatment 3: According to a recent poll, the majority of British national security experts support a retaliatory strike but the majority of British citizens do not support the strike.
- Treatment 4: According to a recent poll, the majority of British national security experts and the majority of British citizens support a retaliatory strike.

Experiment 1 Questions

[JoGo1] {grid roworder=randomize} How strongly do you agree or disagree with the following statements in response to this hypothetical scenario? Please select one option on each row.
-[JoGo1_1] I support the British Government ordering an airstrike against the group responsible for the original cyber-attack
-[JoGo1_2] An airstrike in response to the cyber-attack would likely deter future cyber-attacks
-[JoGo1_3] An airstrike against the group responsible for the cyber-attack would be ethical
<1> Strongly agree
<2> Somewhat agree
<3> Neither agree nor disagree
<4> Somewhat disagree

<5> Strongly disagree
<6> Don't know

[JoGo2] How comfortable or uncomfortable would you be if your support/opposition toward the airstrike in this hypothetical scenario was reported on social media or in a local newspaper? (Please remember your responses will remain anonymous. This is strictly hypothetical.)
<1> Extremely comfortable
<2> Somewhat comfortable
<3> Neither comfortable nor uncomfortable
<4> Somewhat uncomfortable
<5> Extremely uncomfortable
<6> Don't know

[JoGo3] {scale 0 10 dk=1 dk_text="Don't know"} Below are some pairs of statements that describe different attitudes to national security. Please say in each case where your opinion sits on the scale, where 0 and 10 represent opposite opinions.
-[JoGo3_decision] When the Government faces a foreign policy decision, it is more important to follow expert opinion than public opinion|When the Government faces a foreign policy decision, it is more important to follow public opinion than expert opinion

-[JoGo3_experts] National security experts are extremely trustworthy|National security experts are extremely untrustworthy

-[JoGo3_ethics] Acting in the national interest is more important that acting ethically|Acting ethically is more important than acting in the national interest

[JoGo4] Did you think the foreign country in the hypothetical scenario resembled a real-world country?
<1> Yes, but I didn't have any particular country in mind
<2> Yes, and I had a particular country/countries in mind [JoGo4_open] {open}
<3> No, the hypothetical country did not resemble a real-world country
<4> Don't know

Experiment 2 Vignette, Treatments, and Questions

Note: The vignette for U.S. and U.K. respondents differed in the first sentence, which was meant to establish plausibility of the scenario to follow. To avoid deception, I pulled from existing government documents/statements in each country. The rest of the vignette was identical, with adjustments only made for country context.

U.K. Vignette

In 2018, Britain's National Cyber Security Centre's (NCSC) annual review revealed that it was repelling around 10 attempted cyber-attacks every week, with "hostile states" responsible for most of the attacks.

Imagine that a cyberwarfare group in a hostile country enacts a cyber-attack against the electric grid of a British city. Residents are left without power for 3 days and 15 civilians die as their medical life-saving devices cannot be powered.

The British Government is debating whether to engage in an airstrike targeting the cyberwarfare group's headquarters in the foreign country, but the airstrike would probably result in at least 12 civilian casualties.

U.S. Vignette

In 2018, the U.S. National Cyber Strategy revealed that the United States faces frequent cyber-attacks, with hostile states responsible for many of the attacks.

Imagine that a cyberwarfare group in a hostile country enacts a cyber-attack against the electric grid of an American city. Residents are left without power for 3 days and 15 civilians die as their medical life-saving devices cannot be powered.

The U.S. Government is debating whether to engage in an airstrike targeting the cyberwarfare group's headquarters in the foreign country, but the airstrike would probably result in at least 12 civilian casualties.

Treatments groups received a different concluding paragraph to the vignette:

- Control: [shown no information about expert or public opinion]
- Treatment 1: According to a recent poll, the majority of U.S. citizens oppose a retaliatory strike.
- Treatment 2: According to a recent poll, the majority of U.S. national security experts oppose a retaliatory strike.
- Treatment 3: According to a recent poll, the majority of U.S. citizens and the majority of U.S. national security experts oppose a retaliatory airstrike.

Experiment 2 Post-Treatment Questions

[JoGo1] {grid roworder=randomize} How strongly do you agree or disagree with the following statements in response to this hypothetical scenario:
-[JoGo1_1] I support the U.S. Government ordering an airstrike against the group responsible for the original cyber-attack
-[JoGo1_2] An airstrike in response to the cyber-attack would likely deter future cyber-attacks
-[JoGo1_3] An airstrike against the group responsible for the cyber-attack would be ethical
<1> Strongly agree
<2> Tend to agree
<3> Neither agree nor disagree
<4> Tend to disagree
<5> Strongly disagree
<6> Don't know

[JoGo2] How comfortable or uncomfortable would you be if your support/opposition toward the airstrike in this hypothetical scenario was reported on social media or in a local newspaper? (Please remember your responses will remain anonymous. This is strictly hypothetical.)
<1> Extremely comfortable
<2> Fairly comfortable
<3> Neither comfortable nor uncomfortable
<4> Fairly uncomfortable
<5> Extremely uncomfortable
<6> Don't know

[JoGo3] *{scale 0 10 dk=1 dk_text="Don't know"} Below are some pairs of statements that describe different attitudes to national security. Please say in each case where your opinion sits on the scale, where 0 and 10 represent opposite opinions.*
-[JoGo3_decision] When the Government faces a foreign policy decision, it is more important to follow expert opinion than public opinon|When the Government faces a foreign policy decision, it is more important to follow public opinion than expert opinion
-[JoGo3_experts] National security experts are extremely trustworthy|National security experts are extremely untrustworthy
-[JoGo3_ethics] Acting in the national interest is more important that acting ethically|Acting ethically is more important than acting in the national interest

Appendix: Experiment 1

In both surveys, demographic information was collected by YouGov as part of their omnibus survey. In Tables 4 and 5, I provide contingency tables for support for the retaliatory strike (those who responded with a 4 or 5 on a 5-point scale) by various demographic or ideological splits. "% Supports Strike" indicates the percentage of people in a particular group who responded with a 4 or 5. "% of Supporters" indicates the percentage that supporters in that group comprise of all total supporters. In both samples, a higher percentage of people who were male, older, higher income, and on the political right supported retaliating. These findings are consistent with contingency tables for Experiment 2 (Tables 14 and 15).

Summary Statistics

Summary Statistics Experiment 1 (U.K. Sample)

Table 4: Crosstabs for U.K. Sample

Category	Count	% of Total	% Supports Strike	% of Supporters
All	1529	100%	46.2%	100%
Gender MF				
Male	728	47.6%	52.3%	54%
Female	801	52.4%	40.6%	46%
Age Group				
18 - 30	674	44.1%	36.9%	35.3%
30 - 40	228	14.9%	50.4%	16.3%
40 - 55	467	30.5%	51.4%	34%
55+	160	10.5%	63.7%	14.4%
Ethnicity				
English	1364	89.2%	47.9%	92.6%
Non-English	165	10.8%	31.5%	7.4%
Household Income				
<30k	566	37%	49.5%	39.7%
30-70k	505	33%	43.6%	31.2%
70k+	131	8.6%	42.7%	7.9%
Don't Know or Prefer Not to Say	318	20.8%	46.9%	21.1%
NA	9	0.6%	11.1%	0.1%
Party				
Green	83	5.4%	15.7%	1.8%
Other	347	22.7%	41.8%	20.5%
Labour	348	22.8%	31%	15.3%
Conservative	502	32.8%	67.3%	47.9%
Liberal Democrat	117	7.7%	31.6%	5.2%
Scottish National Party (SNP)	52	3.4%	34.6%	2.5%
Brexit Party	26	1.7%	80.8%	3%
UK Independence Party (UKIP)	22	1.4%	77.3%	2.4%
Plaid Cymru	9	0.6%	22.2%	0.3%
Women's Equality Party	3	0.2%	0%	0%
NA	20	1.3%	35%	1%
Voted				
Remain	681	44.5%	33.8%	32.6%
Leave	637	41.7%	64.7%	58.4%
Neither	211	13.8%	30.3%	9.1%
University				
Attended	124	8.1%	25.8%	4.5%
Did Not Attend	1405	91.9%	48%	95.5%
Treatment Group				
Control	295	19.3%	45.1%	18.8%
Both Oppose	333	21.8%	39.6%	18.7%
Experts Oppose, Public Support	285	18.6%	48.1%	19.4%
Experts Support, Public Oppose	279	18.2%	44.4%	17.6%
Both Support	337	22%	53.4%	25.5%

Summary Statistics Experiment 1 (U.S. Sample)

Table 5: Crosstabs for U.S. Sample

Category	Count	% of Total	% Supports Strike	% of Supporters
All	1143	100%	50.8%	100%
Gender MF				
Male	558	48.8%	57.7%	55.4%
Female	585	51.2%	44.3%	44.6%
Age Group				
18 - 30	249	21.8%	34.9%	15%
30 - 40	193	16.9%	43%	14.3%
40 - 55	267	23.4%	52.8%	24.3%
55+	434	38%	62.2%	46.5%
White NonWhite				
Non-White	362	31.7%	46.1%	28.7%
White	781	68.3%	53%	71.3%
Household Income				
<40k	364	31.8%	48.4%	30.3%
40 - 80k	313	27.4%	49.8%	26.9%
80k+	295	25.8%	55.3%	28.1%
Don't Know or Prefer Not to Say	171	15%	50.3%	14.8%
Party Affiliation				
Republican	319	27.9%	73.4%	40.3%
Democratic	426	37.3%	37.3%	27.4%
Independent	310	27.1%	49%	26.2%
Other/Not Sure	88	7.7%	40.9%	6.2%
Supported				
Trump	327	28.6%	78.3%	44.1%
Clinton	349	30.5%	38.1%	22.9%
Neither	467	40.9%	41.1%	33%
University				
Attended	356	31.1%	49.2%	30.1%
Did Not Attend	787	68.9%	51.6%	69.9%
Treatment Group				
Control	236	20.6%	53.4%	21.7%
Both Oppose	225	19.7%	45.3%	17.6%
Experts Oppose, Public Support	231	20.2%	43.7%	17.4%
Experts Support, Public Oppose	209	18.3%	54.1%	19.4%
Both Support	242	21.2%	57.4%	23.9%

Regressions and Robustness Checks

In Tables 6 and 7, I show regression models for the effect of the treatment (as two independent factors with an interaction) on respondent's support for the use of force, whether the respondent believes a retaliation would deter future cyber-attacks, and the extent to which the respondent believes a retaliatory strike would be ethical. Expert opposition has a significant effect (β=-.255 on a five-point scale, p<.05) on perceived effectiveness in the U.K. sample (Table 6) but not on perceived effectiveness in the U.S. sample (Table 7) or on perceived ethics (in either sample). Public opposition does not have a significant effect on perceived effectiveness or on perceived ethics in either sample as well.

Regression Model on Full DVs (U.K. Sample)

Table 6: Regression of Treatment on Support, Effectiveness, and Ethics of Strike (U.K. Sample)

	Dependent variable:		
	Support Strike	Effectiveness Strike	Believe Strike is Ethical
	(1)	(2)	(3)
Experts Oppose	-0.218^{**}	-0.255^{**}	-0.110
	$p = 0.034$	$p = 0.013$	$p = 0.270$
Public Oppose	-0.250^{**}	-0.054	-0.002
	$p = 0.016$	$p = 0.599$	$p = 0.988$
Age	0.009^{***}	0.006^{***}	0.012^{***}
	$p = 0.0001$	$p = 0.006$	$p = 0.00000$
Gender	-0.356^{***}	-0.317^{***}	-0.443^{***}
	$p = 0.00001$	$p = 0.00002$	$p = 0.000$
College	-0.053	-0.0002	-0.051
	$p = 0.470$	$p = 0.998$	$p = 0.477$
Labour	-1.074^{***}	-0.827^{***}	-1.077^{***}
	$p = 0.000$	$p = 0.000$	$p = 0.000$
LibDem	-0.875^{***}	-0.903^{***}	-0.819^{***}
	$p = 0.000$	$p = 0.000$	$p = 0.000$
Other	-0.699^{***}	-0.506^{***}	-0.672^{***}
	$p = 0.000$	$p = 0.00000$	$p = 0.000$
ExpertOppose* PublicOppose	0.067	0.118	-0.068
	$p = 0.644$	$p = 0.414$	$p = 0.629$
Constant	3.806^{***}	3.430^{***}	3.380^{***}
	$p = 0.000$	$p = 0.000$	$p = 0.000$
Observations	1,219	1,177	1,186
R^2	0.164	0.112	0.189
Adjusted R^2	0.157	0.105	0.183
Residual Std. Error	1.263 (df = 1209)	1.236 (df = 1167)	1.211 (df = 1176)
F Statistic	26.264^{***} (df = 9; 1209)	16.285^{***} (df = 9; 1167)	30.400^{***} (df = 9; 1176)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$		

Regression Model on Full DVs (U.S. Sample)

Table 7: Regression of Treatment on Support, Effectiveness, and Ethics of Strike (U.S. Sample)

	Dependent variable:		
	Support Strike	Effectiveness Strike	Believe Strike is Ethical
	(1)	(2)	(3)
Expert Oppose	-0.375^{***}	-0.141	-0.200
	$p = 0.003$	$p = 0.239$	$p = 0.107$
Public Oppose	-0.195	-0.002	-0.055
	$p = 0.128$	$p = 0.986$	$p = 0.666$
Age	0.014^{***}	0.010^{***}	0.015^{***}
	$p = 0.00000$	$p = 0.0002$	$p = 0.00000$
Gender	-0.314^{***}	-0.270^{***}	-0.286^{***}
	$p = 0.001$	$p = 0.003$	$p = 0.002$
College	-0.152	-0.138	-0.004
	$p = 0.115$	$p = 0.138$	$p = 0.969$
Republican	1.151^{***}	1.068^{***}	1.110^{***}
	$p = 0.000$	$p = 0.000$	$p = 0.000$
Independent	0.395^{***}	0.226^{**}	0.345^{***}
	$p = 0.0005$	$p = 0.039$	$p = 0.003$
Other	0.179	-0.247	0.020
	$p = 0.486$	$p = 0.316$	$p = 0.936$
ExpertOppose* PublicOppose	0.195	0.009	0.022
	$p = 0.278$	$p = 0.959$	$p = 0.902$
Constant	2.647^{***}	2.654^{***}	2.341^{***}
	$p = 0.000$	$p = 0.000$	$p = 0.000$
Observations	870	847	850
R^2	0.182	0.162	0.171
Adjusted R^2	0.174	0.153	0.162
Residual Std. Error	1.315 (df = 860)	1.259 (df = 837)	1.304 (df = 840)
F Statistic	21.320^{***} (df = 9; 860)	17.923^{***} (df = 9; 837)	19.213^{***} (df = 9; 840)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$		

Robustness Check of 2x2 Analysis (With and Without Post-Treatment Covariates) (U.K. Sample)

After the treatment and dependent variable questions, I asked respondents the following:

Below are some pairs of statements that describe different attitudes to national security. Please say in each case where your opinion sits on the scale, where 0 and 10 represent opposite opinions.

- When the Government faces a foreign policy decision, it is more important to follow expert opinion than public opinion|When the Government faces a foreign policy decision, it is more important to follow public opinion than expert opinion

- National security experts are extremely trustworthy|National security experts are extremely untrustworthy
- Acting in the national interest is more important that acting ethically|Acting ethically is more important than acting in the national interest

As a robustness check, in Tables 8 and 9 I show the effect of treatments, with and without post-treatment covariates, on respondent support. Although I exclude these questions from the main text analysis because they are influenced by the treatment, regression results across both samples are robust to their inclusion.

Table 8: Regression of Treatment on Support for Strike With and Without Post-Treatment Covariates (U.K. Sample)

	Dependent variable:			
	Support for Strike			
	No Post-Treatment Covs	With Post-Treatment Covs	No Post-Treatment Covs	With Post-Treatment Covs
	(1)	(2)	(3)	(4)
Expert Oppose	-0.215^{*}	-0.214^{*}	-0.237^{**}	-0.227^{**}
	p = 0.083	p = 0.059	p = 0.040	p = 0.037
Public Oppose	-0.214^{*}	-0.258^{**}	-0.223^{*}	-0.259^{**}
	p = 0.089	p = 0.025	p = 0.057	p = 0.019
Follow Experts		0.070^{***}		0.061^{***}
		p = 0.0001		p = 0.0004
Experts Trustworthy		-0.219^{***}		-0.174^{***}
		p = 0.000		p = 0.000
National Interest		0.021		0.019
		p = 0.332		p = 0.370
Age			0.007^{***}	0.005^{**}
			p = 0.007	p = 0.043
Gender			-0.363^{***}	-0.279^{***}
			p = 0.00002	p = 0.0005
College			−0.059	0.052
			p = 0.480	p = 0.521
Labour			-1.095^{***}	-0.825^{***}
			p = 0.000	p = 0.000
LibDem			-0.953^{***}	-0.652^{***}
			p = 0.000	p = 0.00005
Other			-0.686^{***}	-0.492^{***}
			p = 0.000	p = 0.00000
Expert Oppose*, Public Oppose	0.011	0.065	0.036	0.073
	p = 0.953	p = 0.689	p = 0.827	p = 0.640
Constant	3.394^{***}	4.279^{***}	3.917^{***}	4.407^{***}
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Observations	954	954	947	947
R^2	0.012	0.185	0.160	0.257
Adjusted R^2	0.009	0.180	0.152	0.248
Residual Std. Error	1.370 (df = 950)	1.246 (df = 947)	1.268 (df = 937)	1.194 (df = 934)
F Statistic	3.954^{***} (df = 3; 950)	35.940^{***} (df = 6; 947)	19.867^{***} (df = 9; 937)	26.977^{***} (df = 12; 934)
Note:	*p<0.1; **p<0.05; ***p<0.01			

Robustness Check of 2x2 Analysis (With and Without Post-Treatment Covariates) (U.S. Sample)

Table 9: Regression of Treatment on Support for Strike With and Without Post-Treatment Covariates (U.S. Sample)

	Dependent variable:			
	Support for Strike			
	No Post-Treatment Covs	With Post-Treatment Covs	No Post-Treatment Covs	With Post-Treatment Covs
	(1)	(2)	(3)	(4)
Experts Oppose	-0.393^{***}	-0.397^{***}	-0.375^{***}	-0.382^{***}
	p = 0.003	p = 0.002	p = 0.003	p = 0.002
Public Oppose	−0.182	−0.179	−0.195	−0.195
	p = 0.178	p = 0.169	p = 0.128	p = 0.118
Follow Experts			0.395^{***}	0.322^{***}
			p = 0.0005	p = 0.004
Experts Trustworthy			0.179	0.128
			p = 0.486	p = 0.611
National Interest			1.151^{***}	1.031^{***}
			p = 0.000	p = 0.000
Age		0.045^{***}		0.044^{***}
		p = 0.005		p = 0.004
Gender		-0.143^{***}		-0.109^{***}
		p = 0.000		p = 0.000
College		0.025		0.019
		p = 0.145		p = 0.253
Republican			−0.152	−0.118
			p = 0.115	p = 0.213
Independent			0.014^{***}	0.013^{***}
			p = 0.00000	p = 0.00001
Other			-0.314^{***}	-0.279^{***}
			p = 0.001	p = 0.002
Expert Oppose*, Public Oppose	0.169	0.154	0.195	0.190
	p = 0.374	p = 0.403	p = 0.278	p = 0.278
Constant	3.574^{***}	4.197^{***}	2.647^{***}	3.159^{***}
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Observations	907	907	870	870
R^2	0.014	0.093	0.182	0.228
Adjusted R^2	0.011	0.087	0.174	0.217
Residual Std. Error	1.429 (df = 903)	1.372 (df = 900)	1.315 (df = 860)	1.280 (df = 857)
F Statistic	4.275^{***} (df = 3; 903)	15.391^{***} (df = 6; 900)	21.320^{***} (df = 9; 860)	21.073^{***} (df = 12; 857)

Note: *p<0.1; **p<0.05; ***p<0.01

Robustness Check of 2x2 Analysis Without Interaction Term

In Table 1 in the main paper text, I presented models including expert and public opposition as independent instruments. Here, I reproduce the results from Table 1 without the interaction term. The results are consistent with the main text.

Table 10: Regression of Treatment on Support for Strike Without Interaction Term

	Dependent variable:			
	Support for Strike			
	UK Sample (No Covs)	UK Sample (With Covs)	US Sample (No Covs)	US Sample (With Covs)
	(1)	(2)	(3)	(4)
Expert Oppose	-0.186^{**}	-0.185^{**}	-0.312^{***}	-0.283^{***}
	p = 0.018	p = 0.012	p = 0.002	p = 0.002
Public Oppose	-0.230^{***}	-0.216^{***}	-0.097	-0.096
	p = 0.004	p = 0.004	p = 0.309	p = 0.284
Age		0.009^{***}		0.014^{***}
		p = 0.0001		p = 0.00000
Gender		-0.355^{***}		-0.312^{***}
		p = 0.00001		p = 0.001
College		-0.053		-0.151
		p = 0.469		p = 0.119
Labour (UK)		-1.074^{***}		
		p = 0.000		
LibDem (UK)		-0.875^{***}		
		p = 0.000		
Other (UK)		-0.700^{***}		
		p = 0.000		
Republican (US)				1.147^{***}
				p = 0.000
Independent (US)				0.399^{***}
				p = 0.0004
Other (US)				0.176
				p = 0.493
Constant	3.306^{***}	3.790^{***}	3.535^{***}	2.593^{***}
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Observations	1,234	1,219	907	870
R^2	0.013	0.163	0.013	0.181
Adjusted R^2	0.011	0.158	0.011	0.174
Residual Std. Error	1.366 (df = 1231)	1.263 (df = 1210)	1.428 (df = 904)	1.315 (df = 861)
F Statistic	7.866^{***} (df = 2; 1231)	29.540^{***} (df = 8; 1210)	6.018^{***} (df = 2; 904)	23.832^{***} (df = 8; 861)
Note:	*p<0.1; **p<0.05; ***p<0.01			

Regression Models on Comfort Sharing Response

To test whether public cues operate through a pathway of normative influence, I asked respondents:

- How comfortable or uncomfortable would you be if your support/opposition toward the airstrike in this hypothetical scenario was reported on social media or in a local

newspaper? (Please remember your responses will remain anonymous. This is strictly hypothetical.)

In Table 11 and 12, I separated respondents by whether they supported the strike to determine if those who diverge from the public cue are less comfortable sharing their response. Contrary to expectation, in the U.K. sample those who supported the strike were always more comfortable sharing their response than those in the referent category (who responded in the middle of a five-point scale). In the U.S. sample, those who supported the strike were more comfortable sharing their response in four of the five treatments. Only when the experts supported the strike but the public opposed it were "supporters" not significantly more comfortable sharing their response than the referent (β=.331 on a five-point scale, p=.220).

Treatment on Comfort Sharing Response (U.K. Sample)

Table 11: Regression of Treatment on Comfort Sharing Response (U.K. Sample)

	Dependent Variable: Comfort Sharing Response				
	Control	Both Oppose	Expert Oppose, Public Support	Expert Support, Public Oppose	Both Support
	(1)	(2)	(3)	(4)	(5)
Support Strike	0.418*	0.683***	0.682***	0.819***	0.087
	p = 0.075	p = 0.001	p = 0.006	p = 0.0004	p = 0.680
Oppose Strike	−0.175	0.035	0.289	0.118	−0.122
	p = 0.458	p = 0.863	p = 0.249	p = 0.615	p = 0.592
Age	0.001	−0.006	−0.009*	−0.003	−0.005
	p = 0.795	p = 0.161	p = 0.079	p = 0.538	p = 0.257
Gender	−0.322**	−0.427***	−0.367**	−0.380**	−0.451***
	p = 0.035	p = 0.003	p = 0.021	p = 0.015	p = 0.002
College	0.023	−0.103	0.017	−0.014	−0.101
	p = 0.885	p = 0.445	p = 0.917	p = 0.928	p = 0.492
Labour	−0.104	−0.002	−0.097	0.098	−0.005
	p = 0.632	p = 0.991	p = 0.665	p = 0.667	p = 0.983
LibDem	−0.876**	−0.271	0.473	−0.032	−0.631**
	p = 0.015	p = 0.283	p = 0.172	p = 0.914	p = 0.042
Other	0.080	0.014	−0.214	0.037	0.167
	p = 0.674	p = 0.936	p = 0.274	p = 0.853	p = 0.362
Constant	2.851***	3.144***	3.177***	2.866***	3.287***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Observations	264	293	259	262	315
R^2	0.108	0.120	0.082	0.109	0.063
Adjusted R^2	0.080	0.095	0.053	0.080	0.039

Note: *p<0.1; **p<0.05; ***p<0.01

Treatment on Comfort Sharing Response (U.S. Sample)

Table 12: Regression of Treatment on Comfort Sharing Response (U.S. Sample)

	Dependent Variable: Comfort Sharing Response				
	Control	Both Oppose	Expert Oppose, Public Support	Expert Support, Public Oppose	Both Support
	(1)	(2)	(3)	(4)	(5)
Support Strike	0.664***	0.889***	0.726***	0.331	0.462*
	p = 0.002	p = 0.002	p = 0.004	p = 0.220	p = 0.058
Oppose Strike	0.026	0.308	0.137	−0.268	−0.129
	p = 0.911	p = 0.281	p = 0.583	p = 0.370	p = 0.642
Age	−0.003	−0.001	−0.002	0.0002	−0.002
	p = 0.516	p = 0.870	p = 0.693	p = 0.969	p = 0.765
Gender	−0.671***	−0.327*	−0.406**	−0.637***	−0.309*
	p = 0.00002	p = 0.085	p = 0.025	p = 0.001	p = 0.088
College	0.064	0.091	0.122	−0.089	0.029
	p = 0.681	p = 0.643	p = 0.529	p = 0.651	p = 0.882
Republican	0.080	−0.087	0.174	0.273	0.443*
	p = 0.680	p = 0.722	p = 0.466	p = 0.229	p = 0.055
Independent	−0.515***	−0.124	−0.157	−0.298	0.329
	p = 0.004	p = 0.603	p = 0.487	p = 0.221	p = 0.128
Other	−0.109	−0.048	0.378	0.297	0.232
	p = 0.751	p = 0.922	p = 0.515	p = 0.555	p = 0.680
Constant	3.682***	3.012***	3.186***	3.584***	2.977***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Observations	208	196	200	182	217
R^2	0.241	0.097	0.114	0.144	0.095
Adjusted R^2	0.211	0.059	0.077	0.105	0.060
Note:	*p<0.1; **p<0.05; ***p<0.01				

Are Responses Driven by Priors?

One potential concern with the experiment could be that responses to the vignette are driven by respondents' priors. I intentionally do not name a perpetrator country to minimize the likelihood that responses would be driven by prior attitudes toward a certain country. A related concern is that people do not believe the scenario resembles real life, and thus their responses are not likely to be applicable to how their opinions were form if the vignette were to materialize. By not naming a specific country, I run the risk of heightening the hypothesized component of the experiment.

To test whether responses to the vignette are driven by perceptions of a particular country and to probe whether respondents believed the scenario was plausible, I asked respondents in Experiment 1:

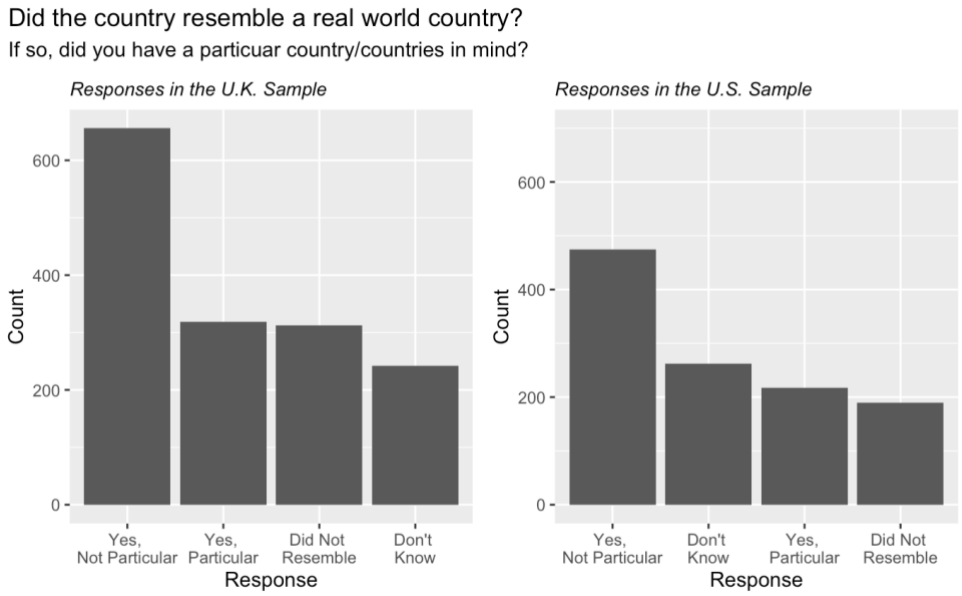
Did you think the foreign country in the hypothetical scenario resembled a real-world country?

- Yes, but I didn't have any particular country in mind
- Yes, and I had a particular country/countries in mind [open response]
- No, the hypothetical country did not resemble a real-world country
- Don't know

As shown in Figure 10, in both the U.K. and U.S. samples, most respondents in both samples said that the foreign country resembled a real-world country, but they did not have a particular country in mind. Among those who said they did have a particular country/countries in mind, most respondents named multiple countries (e.g. "China, Russia, North Korea, Iran"), indicating that their reading of the vignette was unlikely to have been driven by their priors about a single country. Furthermore, Tables 13 and 14 show that the percent of respondents who fall into each category is consistent across treatment groups, suggesting that perception of the country of origin of the attack was not driven by the treatment.

Overall Breakdown of Perception of Perpetrator by Country

Figure 10: Breakdown of Perception of Perpetrator by Sample



Perception of Perpetrator by Treatment Group

Table 13: Breakdown of responses to "Did the country resemble a real world country?," by Treatment Assignment (U.K. Sample)

	1	2	3	4	5
Yes, not particular	42.37	43.11	41.56	38.76	41.32
Yes, particular	19.49	20.89	18.61	19.14	16.94
No	16.53	16	16.02	17.22	17.36
Don't Know	21.61	20	23.81	24.88	24.38
Total	100	100	100	100	100

Table 14: Breakdown of responses to "Did the country resemble a real world country?," by Treatment Assignment (U.S. Sample)

	1	2	3	4	5
Yes, not particular	42.37	43.11	41.56	38.76	41.32
Yes, particular	19.49	20.89	18.61	19.14	16.94
No	16.53	16	16.02	17.22	17.36
Don't Know	21.61	20	23.81	24.88	24.38
Total	100	100	100	100	100

Appendix: Experiment 2

Summary Statistics

Summary Statistics Experiment 2 (U.K. Sample)

Table 15: Crosstabs U.K. Sample

Category	Count	% of Total	% Supports Strike	% of Supporters
All	1578	100%	39.1%	100%
Gender MF				
Male	743	47.1%	48.6%	58.5%
Female	835	52.9%	30.7%	41.5%
Age Group				
18 - 30	259	16.4%	26.6%	11.2%
30 - 40	250	15.8%	28.8%	11.7%
40 - 55	410	26%	39.5%	26.3%
55+	659	41.8%	47.6%	50.9%
Home Income				
<30k	554	35.1%	40.8%	36.6%
30k-70k	513	32.5%	38.4%	31.9%
>70k	130	8.2%	41.5%	8.8%
Don't Know or Prefer Not to Say	352	22.3%	37.8%	21.6%
NA	29	1.8%	24.1%	1.1%
party				
Conservative	465	29.5%	56.6%	42.6%
Labour	474	30%	28.5%	21.9%
LibDem	124	7.9%	29%	5.8%
OtherUK	389	24.7%	38%	24%
NA	126	8%	27.8%	5.7%
Location				
Urban	1206	76.4%	39.1%	76.5%
Rural	207	13.1%	41.1%	13.8%
Town and Fringe	165	10.5%	36.4%	9.7%
Voted				
Remain	714	45.2%	26.5%	30.6%
Leave	659	41.8%	55.1%	58.8%
Neither	205	13%	31.7%	10.5%
University				
Did Not Attend	926	58.7%	42%	63%
Attended	652	41.3%	35%	37%
Treatment Group				
Control	417	26.4%	43.9%	29.7%
Public Oppose	421	26.7%	39.7%	27.1%
Experts Oppose	396	25.1%	37.6%	24.1%
Public and Experts Oppose	344	21.8%	34.3%	19.1%

Summary Statistics Experiment 2 (U.S. Sample)

Table 16: Crosstabs U.S. Sample

Category	Count	% of Total	% Supports Strike	% of Supporters
All	1141	100%	51%	100%
Gender MF				
Male	548	48%	56.6%	53.3%
Female	593	52%	45.9%	46.7%
Age Group				
18 - 30	235	20.6%	35.7%	14.4%
30 - 40	203	17.8%	38.4%	13.4%
40 - 55	261	22.9%	55.6%	24.9%
55+	442	38.7%	62.2%	47.3%
White NonWhite				
White	785	68.8%	53.6%	72.3%
Non-White	356	31.2%	45.2%	27.7%
Household Income				
<40k	376	33%	48.1%	31.1%
40 - 80k	333	29.2%	46.5%	26.6%
80k+	281	24.6%	58.4%	28.2%
Don't Know or Prefer Not to Say	151	13.2%	54.3%	14.1%
Party Affiliation				
Republican	287	25.2%	76%	37.5%
Democratic	417	36.5%	38.6%	27.7%
Independent	330	28.9%	50%	28.4%
Other/Not Sure	107	9.4%	35.5%	6.5%
Supported				
Trump	339	29.7%	77.9%	45.4%
Clinton	391	34.3%	38.1%	25.6%
Neither	411	36%	41.1%	29%
University				
Attended	351	30.8%	48.1%	29%
Did Not Attend	790	69.2%	52.3%	71%
Treatment Group				
Control	265	23.2%	58.5%	26.6%
Public Oppose	309	27.1%	48.5%	25.8%
Experts Oppose	300	26.3%	50%	25.8%
Public and Experts Oppose	267	23.4%	47.6%	21.8%

Regressions and Robustness Checks

In Tables 17 and 18, I show regression models for the effect of the treatment (as two independent factors with an interaction) on respondent's support for the use of force, whether the respondent believes a retaliation would deter future cyber-attacks, and the extent to which the respondent believes a retaliatory strike would be ethical. Expert opposition has a significant effect in all three models in the U.K. sample (Table 17), while it does not have a significant effect in any of the models in the U.S. sample (Table 18). Public opposition has a significant effect on the support and effectiveness models in the U.K. sample (Table 17) and on support and ethics models in the U.S. sample (Table 18).

Regression Model on Full DVs (U.K. Sample)

Table 17: Regression of Treatment on Support, Effectiveness, Ethics (U.K. Sample)

	Dependent variable:		
	Support Strike	Effectiveness Strike	Believe Strike is Ethical
	(1)	(2)	(3)
Public Oppose	-0.199^{**}	-0.197^{**}	-0.140
	$p = 0.029$	$p = 0.028$	$p = 0.116$
Experts Oppose	-0.318^{***}	-0.255^{***}	-0.193^{**}
	$p = 0.001$	$p = 0.005$	$p = 0.033$
Both Oppose	-0.352^{***}	-0.240^{**}	-0.267^{***}
	$p = 0.0003$	$p = 0.011$	$p = 0.005$
Age	0.007^{***}	0.006^{***}	0.009^{***}
	$p = 0.0002$	$p = 0.003$	$p = 0.00002$
Gender	0.470^{***}	0.337^{***}	0.534^{***}
	$p = 0.000$	$p = 0.00000$	$p = 0.000$
College	-0.277^{***}	-0.247^{***}	-0.244^{***}
	$p = 0.00005$	$p = 0.0002$	$p = 0.0003$
Labour	-0.766^{***}	-0.561^{***}	-0.641^{***}
	$p = 0.000$	$p = 0.000$	$p = 0.000$
LibDem	-0.738^{***}	-0.515^{***}	-0.665^{***}
	$p = 0.000$	$p = 0.0001$	$p = 0.00000$
OtherUK	-0.444^{***}	-0.353^{***}	-0.427^{***}
	$p = 0.00000$	$p = 0.00005$	$p = 0.00000$
Constant	3.110^{***}	2.990^{***}	2.757^{***}
	$p = 0.000$	$p = 0.000$	$p = 0.000$
Observations	1,452	1,421	1,421
R^2	0.137	0.087	0.133
Adjusted R^2	0.131	0.081	0.128
Residual Std. Error	1.250 (df = 1442)	1.215 (df = 1411)	1.218 (df = 1411)
F Statistic	25.354^{***} (df = 9; 1442)	14.879^{***} (df = 9; 1411)	24.110^{***} (df = 9; 1411)
Note:	*p<0.1; **p<0.05; ***p<0.01		

Regression Model on Full DVs (U.S. Sample)

Table 18: Regression of Treatment on Support, Effectiveness, Ethics (U.S. Sample)

	Dependent variable:		
	Support Strike	Effectiveness Strike	Believe Strike is Ethical
	(1)	(2)	(3)
Public Oppose	-0.328^{***}	-0.133	-0.315^{***}
	p = 0.003	p = 0.235	p = 0.003
Experts Oppose	-0.165	-0.096	-0.166
	p = 0.126	p = 0.396	p = 0.121
Both Oppose	-0.227^{**}	-0.130	-0.161
	p = 0.041	p = 0.262	p = 0.143
Age	0.014^{***}	0.009^{***}	0.018^{***}
	p = 0.000	p = 0.0002	p = 0.000
Gender	0.245^{***}	0.151^{*}	0.297^{***}
	p = 0.002	p = 0.060	p = 0.0001
College	-0.211^{***}	-0.162^{*}	-0.065
	p = 0.010	p = 0.059	p = 0.420
Republican	0.997^{***}	0.869^{***}	0.908^{***}
	p = 0.000	p = 0.000	p = 0.000
Independent	0.266^{***}	0.269^{***}	0.279^{***}
	p = 0.006	p = 0.007	p = 0.004
Other	0.095	0.144	0.085
	p = 0.494	p = 0.334	p = 0.541
Constant	2.475^{***}	2.581^{***}	2.149^{***}
	p = 0.000	p = 0.000	p = 0.000
Observations	1,141	1,098	1,118
R^2	0.156	0.095	0.162
Adjusted R^2	0.150	0.087	0.155
Residual Std. Error	1.269 (df = 1131)	1.306 (df = 1088)	1.248 (df = 1108)
F Statistic	23.288^{***} (df = 9; 1131)	12.650^{***} (df = 9; 1088)	23.739^{***} (df = 9; 1108)
Note:	*p<0.1; **p<0.05; ***p<0.01		

As a robustness check, in Tables 19 and 20 I show the effect of treatments, with and without post-treatment covariates, on respondent support. Although I exclude these questions from the main text analysis because they are influenced by the treatment, regression results across both samples are robust to their inclusion.

Robustness of 2x2 Analysis (With and Without Post-Treatment Covariates) U.K. Sample

Table 19: Regression of Treatment on Support for Strike With and Without Post-Treatment Covariates (U.K. Sample)

	Dependent variable:			
	Support for Strike			
	No Post-Treatment Covs	With Post-Treatment Covs	No Post-Treatment Covs	With Post-Treatment Covs
	(1)	(2)	(3)	(4)
Public Oppose	−0.149	−0.173*	−0.157	−0.177*
	p = 0.152	p = 0.068	p = 0.121	p = 0.063
Experts Oppose	−0.224**	−0.217**	−0.302***	−0.294***
	p = 0.033	p = 0.024	p = 0.004	p = 0.003
Both Oppose	−0.246**	−0.289***	−0.303***	−0.345***
	p = 0.024	p = 0.004	p = 0.005	p = 0.001
Follow Experts		0.087***		0.085***
		p = 0.00000		p = 0.00000
Experts Trustworthy		−0.205***		−0.165***
		p = 0.000		p = 0.000
National Interest		0.025		0.001
		p = 0.177		p = 0.977
Age			0.007***	0.003
			p = 0.004	p = 0.119
Gender			0.472***	0.401***
			p = 0.000	p = 0.000
College			−0.316***	−0.192***
			p = 0.00003	p = 0.008
Labour			−0.781***	−0.578***
			p = 0.000	p = 0.000
LibDem			−0.752***	−0.464***
			p = 0.00000	p = 0.001
Other			−0.410***	−0.289***
			p = 0.00004	p = 0.002
Expert Oppose*, Public Oppose	3.128***	3.746***	3.177***	3.718***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Observations	1,267	1,267	1,178	1,178
R^2	0.005	0.178	0.142	0.248
Adjusted R^2	0.003	0.174	0.135	0.240
Residual Std. Error	1.338 (df = 1263)	1.218 (df = 1260)	1.253 (df = 1168)	1.175 (df = 1165)
F Statistic	2.188* (df = 3; 1263)	45.347*** (df = 6; 1260)	21.470*** (df = 9; 1168)	31.977*** (df = 12; 1165)
Note:	*p<0.1; **p<0.05; ***p<0.01			

Robustness of 2x2 Analysis (With and Without Post-Treatment Covariates) U.S. Sample

Table 20: Regression of Treatment on Support for Strike With and Without Post-Treatment Covariates (U.S. Sample)

	Dependent variable:			
	Support for Strike			
	No Post-Treatment Covs	With Post-Treatment Covs	No Post-Treatment Covs	With Post-Treatment Covs
	(1)	(2)	(3)	(4)
Public Oppose	-0.327^{**}	-0.337^{***}	-0.322^{***}	-0.332^{***}
	p = 0.012	p = 0.007	p = 0.008	p = 0.005
Experts Oppose	−0.167	−0.165	−0.131	−0.139
	p = 0.204	p = 0.190	p = 0.283	p = 0.243
Both Oppose	-0.259^{*}	-0.242^{*}	-0.217^{*}	-0.211^{*}
	p = 0.056	p = 0.062	p = 0.084	p = 0.085
Follow Experts		0.062^{***}		0.060^{***}
		p = 0.0005		p = 0.0005
Experts Trustworthy		-0.157^{***}		-0.112^{***}
		p = 0.000		p = 0.000
National Interest		0.040^{*}		0.024
		p = 0.051		p = 0.213
Age			0.015^{***}	0.014^{***}
			p = 0.000	p = 0.00000
Gender			0.252^{***}	0.235^{***}
			p = 0.004	p = 0.006
College			-0.225^{**}	−0.145
			p = 0.013	p = 0.104
Labour			0.270^{**}	0.190^{*}
			p = 0.012	p = 0.072
LibDem			0.173	0.058
			p = 0.309	p = 0.729
Other			1.025^{***}	0.882^{***}
			p = 0.000	p = 0.000
Expert Oppose*, Public Oppose	3.588^{***}	4.019^{***}	2.411^{***}	2.799^{***}
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Observations	922	922	922	922
R^2	0.008	0.093	0.161	0.204
Adjusted R^2	0.004	0.087	0.153	0.194
Residual Std. Error	1.400 (df = 918)	1.341 (df = 915)	1.291 (df = 912)	1.259 (df = 909)
F Statistic	2.330^{*} (df = 3; 918)	15.556^{***} (df = 6; 915)	19.433^{***} (df = 9; 912)	19.454^{***} (df = 12; 909)
Note:	*p<0.1; **p<0.05; ***p<0.01			

Regression Models on Comfort Sharing Response

As in Experiment 1, in both the U.S. and U.K. samples in Experiment 2, those who supported the strike were consistently more comfortable sharing their response compared to those who reported that they were neither comfortable nor uncomfortable sharing. This was true across all treatment groups in both countries (see Tables 21 and 22).

Treatment on Comfort Sharing Response (U.K. Sample)

Table 21: Regression of Treatment on Comfort Sharing Response Broken Out by Respondent Support for Strike (U.K. Sample)

	Dependent Variable: Comfort Sharing Response			
	Control	Public Oppose	Experts Oppose	Both Oppose
	(1)	(2)	(3)	(4)
Support Strike	0.818***	0.700***	0.861***	0.440**
	p = 0.00000	p = 0.00004	p = 0.00001	p = 0.016
Oppose Strike	0.105	0.383**	0.405**	−0.176
	p = 0.482	p = 0.023	p = 0.020	p = 0.314
Age	−0.005	−0.001	−0.005	−0.004
	p = 0.160	p = 0.731	p = 0.148	p = 0.322
Gender	0.381***	0.651***	0.323**	0.405***
	p = 0.001	p = 0.00000	p = 0.012	p = 0.004
College	−0.001	0.098	0.073	−0.119
	p = 0.993	p = 0.442	p = 0.581	p = 0.388
Labour	−0.285**	0.018	0.087	−0.060
	p = 0.045	p = 0.908	p = 0.585	p = 0.732
LibDem	−0.216	−0.337	−0.156	−0.283
	p = 0.271	p = 0.215	p = 0.511	p = 0.279
Other	−0.103	0.158	0.438***	−0.151
	p = 0.482	p = 0.323	p = 0.010	p = 0.401
Constant	2.789***	2.357***	2.564***	3.077***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Observations	384	373	359	326
R^2	0.176	0.132	0.119	0.094
Adjusted R^2	0.159	0.113	0.099	0.071
Residual Std. Error	1.050 (df = 375)	1.151 (df = 364)	1.172 (df = 350)	1.184 (df = 317)
F Statistic	10.034*** (df = 8; 375)	6.937*** (df = 8; 364)	5.934*** (df = 8; 350)	4.127*** (df = 8; 317)
Note:	*p<0.1; **p<0.05; ***p<0.01			

Treatment on Comfort Sharing Response (U.S. Sample)

Table 22: Regression of Treatment on Comfort Sharing Response Broken Out by Respondent Support for Strike (U.S. Sample)

	Dependent Variable: Comfort Sharing Response			
	Control	Public Oppose	Experts Oppose	Both Oppose
	(1)	(2)	(3)	(4)
Support Strike	0.880***	0.972***	0.837***	0.693***
	p = 0.00003	p = 0.00001	p = 0.00001	p = 0.001
Oppose Strike	0.030	0.008	0.377*	−0.014
	p = 0.900	p = 0.971	p = 0.076	p = 0.949
Age	−0.010**	−0.009*	−0.006	−0.006
	p = 0.030	p = 0.053	p = 0.209	p = 0.225
Gender	0.237	0.359**	0.338**	0.480***
	p = 0.137	p = 0.018	p = 0.029	p = 0.003
College	−0.160	−0.082	0.028	0.031
	p = 0.353	p = 0.601	p = 0.862	p = 0.851
Republican	0.306	−0.139	0.195	0.323
	p = 0.154	p = 0.479	p = 0.311	p = 0.151
Independent	0.206	0.016	0.078	0.099
	p = 0.283	p = 0.935	p = 0.671	p = 0.613
Other	0.553*	0.339	0.039	0.575**
	p = 0.052	p = 0.207	p = 0.907	p = 0.030
Constant	2.994***	3.066***	2.874***	2.770***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Observations	253	291	287	255
R^2	0.145	0.147	0.119	0.152
Adjusted R^2	0.117	0.123	0.094	0.124
Residual Std. Error	1.202 (df = 244)	1.255 (df = 282)	1.240 (df = 278)	1.187 (df = 246)
F Statistic	5.191*** (df = 8; 244)	6.061*** (df = 8; 282)	4.707*** (df = 8; 278)	5.499*** (df = 8; 246)
Note:	*p<0.1; **p<0.05; ***p<0.01			

Chapter 4: Plausible Deniability and Outsourced Influence Operations

Dissertation Paper 3
Josh A. Goldstein[1]

Abstract

Foreign actors waging propaganda campaigns often take operational security measures to hide their identity—including outsourcing their campaigns to private contractors. The covert action literature suggests that for kinetic operations such plausible deniability has benefits, such as limiting reputational backlash and retaliation from the target. At the same time, recent literature suggests there may be benefits to waging operations that are *implausibly* deniable, such as appearing omnipotent and stirring paranoia in the target state. We know very little about the relative costs and benefits of plausible and implausible deniability for cyber-enabled influence operations, and whether these costs and benefits are equally accessible for all potential perpetrator countries. I deploy a survey experiment that varies whether different foreign governments are implicated in influence operations targeting the 2024 U.S. presidential election campaign in different ways. I find that when governments are not implicated in operating or paying for an influence operation that originates in their country, they can avoid blowback and yet are no less effective at stirring paranoia or appearing omnipotent. I add the novel insight that assignment of blame and support for retaliation heavily depend on the country of origin: the effects of plausible versus implausible deniability are stronger for allies than adversaries, and respondents are much more likely to blame an adversarial government for operations originating on its soil—even when the government's role is not known.[2]

[1] For useful input on the project at various stages, I thank Jonathan Chu, Janina Dill, Shelby Grossman, Kathryn Hedgecock, Lindsay Hundley, Leah Matchett, Lauren Sukin, Mike Tomz, Jessica Weeks, and Scott Williamson.
[2] This project has received ethical approval from the University of Oxford under DPIR Research Ethics Approval: SSH_DPIR_C1A_21_014.

Introduction

In the lead-up to the 2020 U.S. election, CNN broke a major story about foreign interference: unlike in 2016 when Russian influence operations targeting the United States were run from Russia, Russian influence operations in 2020 were, at least in part, outsourced to people in Ghana and Nigeria. Social media accounts run by operators in Ghana and Nigeria pretended to be American and spread polarizing content about racial issues in the United States.[3] As David Imbordino, election threat lead at the U.S. National Security Agency, said:

> In terms of 2020 [in the Internet Research Agency] we've seen a shift towards more use of proxies…intermediaries…laundering information through other individuals in the media space. They had set up something in Africa, in Ghana—in terms of having people there trying to put stuff online about divisive issues, using covert influence websites being able to get their narrative out.[4]

The 2020 case is but the latest example of a rising trend: governments or political groups outsourcing their influence operations to third-party actors. In 2020 alone, Facebook and Twitter attributed at least 15 networks of accounts to private firms, such as marketing or PR companies.[5] This is likely a significant undercount of the number of influence operations run by marketing or PR companies on social media. These are only the networks that Facebook and Twitter have found and chosen to publicly announce; as Facebook's own Head of Security Policy has said, when it comes to influence operations "we should be careful about saying that we know what the denominator is."[6] Moreover, Facebook and Twitter are only two of the many social media companies. Finally, accounts in each network often engage in multiple discrete operations, such as sharing content about multiple elections. Thus, treating the number of takedown networks as a proxy for the number of influence operations significantly understates the latter.[7]

---

[3] Ward, Clarissa, Katie Polglase, Sebastian Shukla, Gianluca Mezzofiore, and Tim Liste. "Russian election meddling is back – via Ghana and Nigeria – in your feeds." *CNN.* April 11, 2020. Accessed: https://www.cnn.com/2020/03/12/world/russia-ghana-troll-farms-2020-ward/index.html

[4] Vavra, Shannon. "Election interference efforts have shifted, NSA and Cyber Command election threats lead say." *CyberScoop.* August 7, 2020. Accessed: https://www.cyberscoop.com/election-interference-2020-presidential-elections-nsa-cyber-comand-russia-china/

[5] Goldstein, Josh A. and Shelby Grossman. "How Disinformation Evolved in 2020." *Brookings TechStream.* January 4, 2021. Accessed: https://www.brookings.edu/techstream/how-disinformation-evolved-in-2020/

[6] Dwoskin, Elizabeth. "Russia is still the biggest player in disinformation, Facebook says." *The Washington Post.* May 26, 2021. Accessed: https://www.washingtonpost.com/technology/2021/05/26/facebook-disinformation-russia-report/

[7] For a historical analogy, agents of influence have often been used by intelligence agencies for multiple influence operations. A network of accounts is akin to an agent of influence that can be deployed for different ends.

Why are outsourced influence operations on the rise? One obvious reason is that as the disinformation research community becomes better at tracking and exposing influence operations, marketing firms offer plausible deniability. A team of influence operations experts from Facebook write that "IO-for-hire" entities could be used "by sophisticated actors to hide their involvement behind private firms, making attribution more challenging."[8] In one of the few pieces to examine "IO-for-hire," sometimes referred to as "outsourced disinformation," Grossman and Ramali suggest that plausible deniability is the primary advantage of outsourcing influence operations:

> The primary advantage [of outsourcing] is that it gives the government a level of plausible deniability. If the operation is uncovered, government actors can claim that it was simply a rogue social media marketing agency, and that they had nothing to do with the activities. Similarly, if one firm gets banned from social platforms, governments can switch to working with a new one.[9]

The conventional wisdom is that states waging influence operations benefit from operational security measures that hide their identity and offer plausible deniability. However, the covert action literature, which typically focuses on kinetic forms of covert action like assassination and regime change, provides reason to believe that plausible deniability is better conceived as a trade-off for the perpetrator of an influence operation rather than as an unequivocal benefit.[10] On the one hand, plausible deniability may allow a state to limit reputational backlash for actions that violate international laws or norms, and to minimize the likelihood of escalation from the target.[11] On the other hand, *implausible* deniability—when a state's denial would be implausible—has its own benefits: it may make the perpetrator appear

[8] Gleicher, Nathaniel, Margarita Franklin, David Agranovich, Ben Nimmo, Olga Belogolova, and Mike Torrey. 2021. "The State of Influence Operations 2017-2020" *Facebook Report.* p. 24. Accessed: https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf

[9] Grossman, Shelby and Khadeja Ramali. 2020. "Outsourcing Disinformation." *Lawfare*. Accessed: https://www.lawfareblog.com/outsourcing-disinformation

[10] For recent examples from the kinetic covert action literature, see: Downes, Alexander B., and Mary Lauren Lilley. 2010. "Overt Peace, Covert War? Covert Intervention and the Democratic Peace." *Security Studies* 19(2): 266–306; Carson 2018; Joseph, Michael F. and Michael Poznansky. 2018. "Media Technology, Covert Action, and the Politics of Exposure." *Journal of Peace Research* 55(3): 320–35.

[11] Carnegie, Allison. 2021. "Secrecy in International Relations and Foreign Policy." *Annual Review of Political Science* 24: 213-233; Carson, Austin. 2016. "Facing Off and Saving Face: Covert Intervention and Escalation Management in the Korean War." *International Organization* 70(1): 103-131; Poznansky, Michael. 2019. "Feigning Compliance: Covert Action and International Law." *International Studies Quarterly.* 63(1): 72-84.

strong and omnipotent to audiences in the target state and stir paranoia there.[12] States thus face a trade-off when preparing influence campaigns. They must determine the extent to which they seek to conceal their operations: whether to pursue the benefits of plausible or implausible deniability.

If plausible deniability is not uncontestably a benefit for the perpetrator, then further work is necessary to determine the relative benefits of plausible versus implausible deniability. These benefits depend, at least in part, on how citizens in the target state would react to different attributions of influence operations. However, limited research to date has explored public reactions to foreign influence operations under conditions of uncertain attribution.[13] No studies to date have systematically unpacked the trade-offs between plausible and implausible deniability as applied to influence operations or tested public responses to in-house versus out-sourced influence operations.[14]

Building on the literatures on foreign interference and covert action, I ask two primary questions: First, how does uncertainty over whether a foreign government operated or paid for a foreign influence operation affect the way that Americans view the operation and the alleged perpetrator? And second, how does uncertainty over the foreign government's involvement influence Americans' support for retaliation?

To assess the benefits of plausible versus implausible deniability, I theoretically unpack the different types of uncertainty that a target public may face over a foreign influence operation and test the effects of a previously unexplored form of uncertainty. I argue that uncertainty can exist over (I) whether an operation took place, (II) whether it was conducted by domestic citizens or foreigners, and, if it was conducted by foreigners, (III) whether it was operated and/or paid for

---

[12] Cormac, Rory and Richard J. Aldrich. 2018. "Grey is the new black: covert action and implausible deniability" *International Affairs* 94(3): 477-494.
Note: I do not study when state's claim responsibility for the operation outright for two reasons. First, it is empirically rare for a government to claim credit for a covert informational campaign in the immediate aftermath of the operation. Second, if states did take credit directly, the operation would, by definition, not be covert. This thus falls outside the scope of plausible versus implausible deniability.

[13] Tomz, Michael and Jessica L. P. Weeks. 2020. "Public Opinion and Foreign Electoral Intervention." *American Political Science Review* 114(3): 856-873; Tomz, Michael and Jessica L. P. Weeks. "Rhetoric, Partisanship, and Public Reactions to Foreign Electoral Intervention." *Working Paper.* (n.d.-a)

[14] For example, if plausible deniability works similarly in both domains, we should expect that if a perpetrator's influence operation is plausibly deniable, the public in the target state will be less likely to support retaliation (Carson 2016). Although it does not focus on public views of government-run versus outsourced disinformation campaigns, one notable working paper considers how the Russian Internet Research Agency's influence operations against the United States differed from those conducted by Russian military intelligence (the GRU). See: DiResta, Grossman, and Siegel n.d.

by a particular foreign actor.[15] I build on existing work by investigating an unexplored form of Type III Uncertainty—when an influence operation can be traced to a foreign country, but the government's role is not known.

I conduct a survey experiment to investigate how Americans respond to foreign influence operations targeting U.S. elections when a campaign is operated or paid for by a foreign government compared to when it is not known whether the government was involved. To deploy the survey, I use Lucid Theorem (n=7,792). The survey presents American respondents with a vignette of an alleged foreign social media influence operation targeting the 2024 U.S. presidential election campaign. I vary whether the operation was conducted by a foreign government, a marketing firm in a foreign country, or an unknown actor, and, if it was conducted by a marketing firm, whether the marketing firm was paid by the foreign government, foreign private citizens, or an unknown entity. Political actors increasingly outsource their disinformation work to private companies,[16] in part for the supposed benefits of plausible deniability.[17] Social media platforms often have data to track an operation to a particular country of origin, but there is variability in whether an operation can be attributed to specific actors therein. The variation in the vignette thus mirrors a real-world type of uncertainty that can create plausible deniability for a government waging influence operations.

I look at three outcomes. First, one benefit of plausible deniability in the literature on kinetic covert action is that it limits public pressures in the target country to escalate.[18] I apply this logic to the informational space and test whether plausible deniability limits public support for retaliation.[19] A second theorized benefit of plausible deniability in the kinetic covert action literature is that it limits reputational damage to the perpetrator.[20] I therefore test whether in reaction to a social media influence operation, people in the target state are less likely to hold negative views of the foreign country when the operation is plausibly deniable compared to when it is attributed to a foreign government (i.e. implausibly deniable). Third, the covert action

---

[15] I thank Jessica Weeks for helping me develop these points.
[16] Bradshaw, Samantha and Philip N. Howard. 2019. "The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation." *Oxford Internet Institute Report;* Martin, Diego A., Jacob N. Shapiro, and Julia G. Ilhardt. 2020. "Trends in Online Influence Efforts." *Working Paper: Version 2.0*. August 5, 2020. Accessed: https://scholar.princeton.edu/sites/default/files/jns/files/trends_in_online_influence_efforts_v2.0_aug_5_2020.pdf
[17] DiResta, Renée, Shelby Grossman, and Alexandra Siegel. n.d. "In-House vs. Outsourced Trolls: How Digital Mercenaries Shape State Influence Strategies." *Working Paper;* Gleicher et al. 2021; Grossman and Ramali 2020.
[18] Carson 2016.
[19] I also relatedly look at how plausible deniability impacts the public's assignment of blame.
[20] Poznansky, Michael. 2020. "Revisiting plausible deniability." *Journal of Strategic Studies*. 1-22.

literature theorizes that when foreign interference is *implausibly* deniable, it makes the perpetrator state appear omnipotent and stirs paranoia.[21] It may be perceived as a sign of strength that a country flouts norms of non-interference, and citizens in the target state may begin to see the perpetrator's hand behind operations or discourse in which it was not involved. I apply this logic by testing whether people are more concerned about the perpetrator's disinformation campaigns and view the country as more influential in American politics when an operation is conducted or paid for by a foreign government compared to when it is not known if the government was involved.[22]

I find small but significant effects for the theorized benefits of plausible deniability. When an operation cannot be confidently attributed to a foreign government, Americans are less likely to support lower-level forms of retaliation and less likely to hold an unfavorable view of the perpetrator country. However, I do not find evidence for the theorized benefits of implausible deniability. When an operation can be confidently attributed to a foreign government, Americans do not view the country as stronger, the operation as more effective, and Americans are no more likely to be paranoid of influence operations from the perpetrator country. These results suggest that the likely calculation of political actors outsourcing influence operations to third-party firms is correct. Political actors can reap the benefits of plausible deniability and do not forego the supposed benefits of getting caught.

I also build on existing work on uncertainty about foreign interference by testing how plausible deniability is impacted by a respondent's prior view of the country from which the operation originated.[23] In my survey, I experimentally manipulate whether the perpetrator is a country with high or low favorability in the United States (Russia and China, compared to the United Kingdom and Japan) to determine how plausible deniability is moderated by prior views of the alleged perpetrator. I find that the identity of the country of origin of the operation has a large and significant effect on whether respondents believe a government was involved or is

[21] Cormac and Aldrich 2018.
[22] As described above, there is a trade-off between the first two logics and the third. If a state achieves plausible deniability in an operation, they can limit escalation and limit reputational damage, but they are less likely to benefit from omnipotence and stir paranoia in the target state than if their operation is implausibly deniable. If a state's operation is implausibly deniable, they can appear more powerful or stir paranoia, but this comes at the potential risk of escalation and reputational damage.
[23] Tomz and Weeks 2020; Tomz and Weeks n.d.-a; Williamson, Scott. n.d. "Proxies and Plausible Deniability" *Working Paper*.

responsible for an operation, and whether respondents support retaliating. I also find the effect of the plausible deniability treatments is much larger for allies than adversaries.

While the findings mentioned above are not surprising, the relative magnitudes of the effects are: the identity of the country of origin had a much larger effect than the level of deniability on whether a respondent viewed a government as responsible for an operation and whether a respondent supported retaliation. The country of origin of an operation is more important in the eyes of the American public than the type of attribution.

These results have significant implications for understanding and combatting foreign influence operations. The fact that the country of origin has such a strong effect may explain why actors who develop a reputation for waging covert influence operations continue to wage operations that are brazen and likely to get caught. Once a reputation is developed, governments are unlikely to benefit from plausible deniability, so they may believe that their focus should be more on operational effectiveness than retaining secrecy or a scapegoat. Second, because outsourcing to third-party actors can yield the benefits of plausible deniability for some perpetrators, government and civil society actors face the challenge of disincentivizing marketing and PR firms from partaking in covert political campaigns. Finally, in response to my vignette, many Americans believed that thousands of social media accounts likely had an effect on election results, even though no information about engagements or impact was provided in the vignette. This represents a vulnerability, since creating fake accounts is both frequent and not prohibitively costly.

Secrecy and Foreign Covert Action

Within the last two decades, declassification of Cold War records has spurred research in International Relations on secrecy and deception.[24] Two strands of this literature are pertinent to the question of how the public perceives foreign influence operations under conditions of uncertainty.

The first relevant contribution of the covert action literature renders problematic the way that secrecy and plausible deniability were treated within International Relations. Scholars have shown that both secrecy and plausible deniability are not binaries, but exist on spectra. For

[24] Myrick, Rachel. 2020. "Why So Secretive? Unpacking Public Attitudes toward Secrecy and Success in US Foreign Policy." *Journal of Politics* 82(3): 828-843.

example, Poznansky and Perkoski bring the distinction between covert and clandestine operations to the fore to highlight that covert action is a political choice.[25] Clandestine operations are defined by the Department of Defense as "operations sponsored or conducted by governmental departments or agencies in such a way as to assure secrecy or concealment."[26] This relies on a tactical or 'operational necessity' logic, whereby secrecy is necessary for the mission to occur or secrecy improves the operation's probability of success. Covert operations, in contrast, are codified in U.S. law in NSC 10/2 as activities "which are so planned and executed that any US Government responsibility for them is not evident to unauthorized persons and that if uncovered the US Government can plausibly disclaim any responsibility for them."[27] Unlike in the case of clandestine operations, states attempt to disclaim responsibility for covert operations *even after they occur*. By disaggregating secrecy, Poznansky and Perkoski show that there are four types of secret action based on whether an operation is executed in secret (clandestine) and whether it is denied by the sponsor afterward (covert).[28] In doing so, they show that covert action is a political choice, not a tactical necessity.[29]

Similarly, Cormac and Aldrich argue that plausible deniability is often a misnomer. Foreign actors are frequently aware of covert operations and the evidence linking the perpetrator to the supposedly "covert" action can be convincing.[30] Carson and Yarhi-Milo show that local partners and strategic adversaries sometimes observe covert action, and that target states find covert communication "both intelligible (the basic intended message is understood by perceivers) and credible (the message is believable)."[31] They analogize covert action to a theatrical stage, where state actors engage in backstage behavior that is mutually intelligible, while concealing some of this behavior from the audience.

The existing literature establishes that plausible deniability and secrecy exist on spectra, and thus that there is variation in the extent to which a foreign public will be aware of covert

---

[25] Poznansky, Michael and Evan Perkoski. 2018. "Rethinking Secrecy in Cyberspace: The Politics of Voluntary Attribution." *Journal of Global Security Studies*. 3(4): 402-416.
[26] Joint Chiefs of Staff, DOD Dictionary of Military and Associated Terms, April 2019, p. 37, at https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/dictionary.pdf
[27] https://history.state.gov/historicaldocuments/frus1964-68v26/actionsstatement; https://history.state.gov/historicaldocuments/frus1945-50Intel/d292
[28] These dimensions (covert and clandestine) form a 2x2 in the Poznansky and Perkoski formulation.
[29] Poznansky and Perkoski 2018.
[30] Cormac and Aldrich 2018.
[31] Carson, Austin and Keren Yarhi-Milo. 2017. "Covert communication: The Intelligibility and Credibility of Signaling in Secret." *Security Studies* 26(1): 124.

action and in the extent to which a foreign government or media will acknowledge operations. However, the existing literature does not investigate how different types of deniability and secrecy are perceived by citizens in the target state. This is a shortcoming because different types and degrees of deniability exist, both in cyber writ large and in relation to cyber-enabled influence operations specifically.

As Rid and Buchanan write about cyber-attacks, "attribution is a nuanced and multi-layer process, not a problem that can simply be solved or not solved."[32] Lin unpacks different layers of attribution in the domain of cyber-attacks: attribution can sometimes be made to a machine, a human behind the machine, or to a party ultimately responsible for the activity.[33] Sometimes it is possible to make attribution at one layer but not another, and, in some cases, it is not possible to make attribution at all.

As in the case of cyber-attacks, uncertainty over the perpetrator of an influence operation—particularly in the immediate aftermath of an operation—is not only plausible, but likely. In the Russian Internet Research Agency's social media operation during the 2016 U.S. election campaign, operational security measures were lacking. For example, Facebook ads about polarizing content targeting Americans were paid for in Russian rubles; some accounts pretending to be Americans used a single IP address in Venezuela or the Russian Internet Research Agency's real IP address.[34] More recently, states running social media influence operations often take additional steps for operational security to conceal their identities. These operational security measures depend on the tactics that researchers trying to expose influence operations use, creating a cat-and-mouse dynamic in the disinformation space.[35] Operational security measures seen in recent influence operations include using AI-generated photos to evade

[32] Rid, Thomas and Ben Buchanan. 2015. "Attributing Cyber Attacks." *Journal of Strategic Studies.* 38(1-2): 30.
[33] Lin, Herbert. 2016. "Attribution of Malicious Cyber Incidents: From Soup to Nuts." *Journal of International Affairs.* 70(1): 75-137.
[34] DiResta, Renée, Kris Shaffer, Becky Ruppel, David Sullivan, Robert Matney, Ryan Fox, Jonathan Albright, and Ben Johnson. 2018. "The Tactics & Tropes of the Internet Research Agency." *New Knowledge Report*. Accessed: https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1003&context=senatedocs; Howard, Philip N., Bharath Ganesh, Dimitra Liotsiou, John Kelly, Camille Francois. 2019. "The IRA, Social Media, and Political Polarization in the United States, 2012-2018." Oxford Internet Institute Paper. Accessed: https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1004&context=senatedocs
[35] Goldstein and Grossman 2021.

reverse-image searching,[36] using VPNs to route computers through the United States,[37] using front non-profits in third-party countries to run propaganda campaigns,[38] and hiring unwitting journalists to write copy for faux news organizations.[39] If states attempt to conceal their identity and attribution is not always clear, further research is needed to understand how the public responds under various conditions of uncertainty.

Existing Theories of Plausible and Implausible Deniability

A second relevant contribution of the covert action literature has been to theorize why states opt for covert, instead of overt, action. This work asks: What are the benefits of plausible deniability?

In an instructive piece dedicated to deciphering the logics of plausible deniability, Poznansky posits two models of secrecy.[40] Poznansky's first model is an Executive Model, where the executive seeks plausible deniability from their domestic public. Governments opt for covert action, here, to protect the reputation of the leader from domestic, not international, audiences. Secrecy is meant to shield the head of state from domestically unpopular acts or acts that would be illegal within the domestic context. I set aside rationales for secrecy under the Executive Model for future research.[41]

Poznansky's second model is more pertinent to the research at hand. This is a State Model, where the state keeps its actions secret or plausibly deniable vis-à-vis a foreign public. States may opt for secrecy under the state model for two reasons.[42] First, they may want to limit escalatory pressures. According to Carson, overt action may lead publics in both countries to

[36] Graphika. 2020. "Operation Naval Gazing: Facebook Takes Down Inauthentic Chinese Network." *Graphika Technical Report*. Accessed: https://public-assets.graphika.com/reports/graphika_report_naval_gazing.pdf
[37] Lee, Dave. "The tactics of a Russian troll farm." *BBC.* February 16, 2018. Accessed: https://www.bbc.com/news/technology-43093390
[38] Ward, Clarissa, Katie Polglase, Sebastian Shukla, Gianluca Mezzofiore, and Tim Lister. "Russian election meddling is back – via Ghana and Nigeria – in your feeds." *CNN*. April 11, 2020. Accessed: https://www.cnn.com/2020/03/12/world/russia-ghana-troll-farms-2020-ward/index.html
[39] Stubbs, Jack. "Exclusive: Russian operation masqueraded as right-wing news site to target U.S. voters – sources." *Reuters*. October 1, 2020. Accessed: https://www.reuters.com/article/usa-election-russia-disinformation/exclusive-russian-operation-masqueraded-as-right-wing-news-site-to-target-u-s-voters-sources-idUSKBN26M5OP
[40] Poznansky 2020. See also: Gibbs, David N. 1995. "Secrecy and International Relations." *Journal of Peace Research* 32(2): 213-228; Carnegie 2021.
[41] See also: Tomz, Michael and Jessica L. P. Weeks. "The Domestic Political Costs of Soliciting Foreign Electoral Intervention." *Working Paper.* (n.d.-b)
[42] Poznansky 2020.

place pressure on decision-makers to escalate.[43] This helps to explain why governments in both the perpetrator and target country sometimes collude in keeping covert intervention secret. The second rationale, to avoid violations of international law and norms, builds on the literature of state reputation. States don't want to lose trust and credibility in the international community, nor face the reputational costs of appearing hypocritical. This drives states that care about their reputation to pursue covert intervention instead of overt intervention when a legal exemption from nonintervention is unavailable.[44]

Poznansky's two rationales for secrecy under the state model focus on plausibly deniable behavior. However, there is also reason to pursue operations that are *implausibly* deniable. Cormac and Aldrich highlight that when states pursue covert action that is implausibly deniable, they reap benefits such as omnipotence: if their actions are revealed but they do not take credit, people in the target country may see the hand of the perpetrator as ever-present in domestic politics.[45] This creates a trade-off between reaping the benefits of plausible deniability (limit escalation and avoid the perception of violating international norms) and the benefits of implausible deniability (convey power and stir paranoia). I summarize these logics in Figure 2.

The logics of plausible and implausible deniability in the covert action literature primarily stem from, and are used to describe, kinetic forms of covert action—such as assassination and regime change. However, limited work has applied these logics to influence operations.

In the following section, I review several relevant experimental studies on uncertainty or plausible deniability. These studies provide mixed findings for the extent to which certainty in attribution and plausible deniability affect the public's views of covert action. They also study subtly different forms of uncertainty. I then move to theory development and make two contributions. First, I provide a framework for thinking about uncertainty vis-à-vis foreign influence operations. This helps to show how I contribute by testing a novel form of uncertainty. Second, I apply the existing theories of plausible and implausible deniability from the kinetic covert action literature to influence operations. I argue that some of the expectations about plausible and implausible deniability theoretically apply, while others do not. For the rationales

[43] Carson 2018.
[44] Poznansky 2019.
[45] Cormac and Aldrich 2018.

that apply, I develop testable hypotheses, which I proceed to investigate through an original survey experiment.

Mixed Experimental Evidence on Uncertainty and Covert Action

Recent studies have begun to probe how uncertainty impacts how the public in a target state views hostile action by a foreign power, coming to mixed conclusions. First, Tomz and Weeks conduct survey experiments to investigate how Americans view different types of foreign electoral interference. Although it is not the main focus of the paper, the authors create uncertainty about the source of interference. "Recognizing that citizens might not be sure which foreign country carried out the intervention, we randomized *percent* to be 50%, 75%, 95%, or 100%."[46] The authors find some evidence that respondents favored retaliating when there were higher degrees of certainty, but the effects were generally small.

Second, two studies addressing public responses to cyber-attacks come to different conclusions about the importance of attribution certainty. Kreps and Das experimentally vary certainty regarding the perpetrator's involvement in the operation, telling respondents that it was "probably" or "almost certainly" involved in the operation. The authors find that the degree of certainty in the attribution "had no appreciable impact, perhaps because respondents are unable to distinguish between 'probably' and 'almost certainly.'"[47] In a second study on public responses to cyber-attacks, Hedgecock et al. vary four features of the experiment: whether an operation is kinetic or cyber, the scale of the operation (whether it results in information theft, financial theft, infrastructure destruction, or loss of life), whether the U.S. government has high or low confidence in its attribution to a hypothetical terrorist group, and the timing of the operation (whether it occurred recently or more than a year ago).[48] Hedgecock et al. find that the largest treatment effect is associated with attribution: almost three-quarters of respondents favor retaliation when the attack is made with "high confidence," compared to under 60% when it is made with "low confidence."

[46] Tomz and Weeks 2020, 7.
[47] Kreps, Sarah and Debak Das. 2017. "Warring from the virtual to the real: Assessing the public's threshold for war over cyber security." *Research and Politics.* p. 4.
[48] Hedgecock, Kathryn, Lauren Sukin, and Leah Matchett. "The Distinctiveness of Public Response to Cyber and Kinetic Attacks." *Working Paper*.

Lastly, Williamson studies the effects of uncertainty over a foreign government's role in an operation.[49] Williamson investigates whether attributions of blame and demands for retaliation vary based on whether a foreign government conducts a covert military operation themselves or uses proxies to carry it out. When an operation was conducted by proxies, respondents were less likely to believe it was carried out on orders of the foreign government. However, respondents were no less likely to support a retaliatory attack based on whether the operation was conducted by the government or foreign proxies.

These studies focus on public opinion about different substantive questions—foreign electoral interference, kinetic versus cyber-attacks, and government-run versus proxy-run kinetic covert action. They also study different types of uncertainty. First, Tomz and Weeks tell respondents that a foreign government has interfered, but create uncertainty over *which* foreign government it is.[50] Second, Kreps and Das create uncertainty over whether Russia was the perpetrator of an influence operation; unlike Tomz and Weeks, they do not first tell respondents that the operation necessarily stemmed from a foreign source or a state actor.[51] Likewise, Hedgecock et al. study uncertainty over whether a particular foreign actor (Red Square, a hypothetical group sponsored by Russia) conducted the operation, and do not first tell respondents that the operation necessarily stemmed from a foreign actor.[52] Finally, Williamson creates uncertainty over whether the foreign government ordered the operation by attributing it to a government directly, a state-run corporation, or a private corporation.[53] While these studies are useful for supporting the notion that plausible deniability in the domain of influence operations may have effects, none provides a framework for comparing different types of uncertainty or focuses primarily on covert informational campaigns.

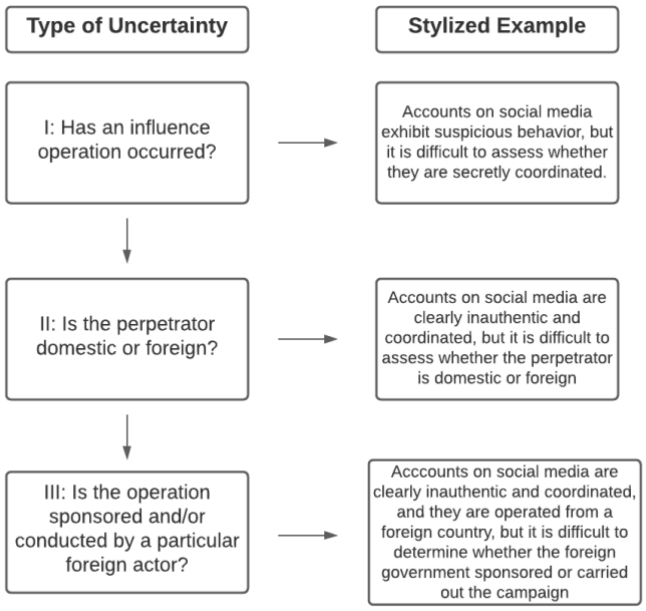
Theorizing Forms of Uncertainty

Building on the different experimental treatments for creating uncertainty that are described above, I provide a framework for thinking about distinct types of uncertainty as applied to foreign influence operations (as shown in Figure 1). First, there can be uncertainty

[49] Williamson n.d.
[50] Tomz and Weeks 2020.
[51] Kreps and Das 2017.
[52] Hedgecock et al. n.d.
[53] Williamson n.d.

over whether a foreign influence operation has occurred at all. For example, if accounts on social media use the same hashtag in unison, it may be difficult to determine whether they are covertly coordinated or exhibiting authentic user behavior. Second, if an operation has occurred, there may be uncertainty over whether the actor that conducted the operation was domestic or foreign. For example, it may be possible to prove that social media accounts are coordinated and inauthentic, but difficult to determine whether they are run by Americans or foreigners. Third, if a foreign operation has occurred, there may be uncertainty over who is responsible for conducting and/or sponsoring the operation. For example, it may be possible to establish that inauthentic social media accounts targeting Americans are operated from a foreign country, but difficult to determine who operates the accounts or who paid for the operation.

Figure 1: Visualization of Three Types of Uncertainty



Each of the three layers of uncertainty warrants further attention. Studying Type I uncertainty could help to explain the conditions under which Americans will believe an influence operation has taken place. In a working paper, Tomz and Weeks study how the public responds to allegations of foreign interference in elections by the U.S. intelligence community when the winning candidate denies that involvement took place.[54] Researchers could study public

[54] Tomz and Weeks n.d.-a.

reactions when a loser alleges that foreign intervention has occurred, or when the media reports a foreign influence operation has taken place but other credible actors—such as candidates—deny this is the case. Type II uncertainty raises questions about whether Americans respond differently to influence operations that are domestic or foreign in origin. A transparency logic would suggest that Americans would disapprove of covert informational campaigns regardless of the source, but social identity theory might imply that Americans will be more sensitive to foreign influence operations.[55] Type III uncertainty raises questions about how Americans respond to foreign influence operations that are state- versus privately- ordered and run. Outsourcing is a way to obtain plausible deniability, and significant research in covert action theorizes that there are both benefits and drawbacks to plausible deniability.

I investigate how Americans respond to foreign influence operations when there is uncertainty over the identity of the foreign actor that operated and/or paid for a particular influence operation. I leave future research to further illuminate Type I Uncertainty, Type II Uncertainty, and alternative forms of Type III Uncertainty. The most relevant literature for developing expectations about plausible and implausible deniability focuses primarily on kinetic forms of covert action, as described above. Thus, in the following section, I examine whether the logics of plausible and implausible deniability from the covert action are likely to apply to non-kinetic, informational forms of covert action. I develop hypotheses for the logics that do apply, which I proceed to test through a survey experiment on Type III Uncertainty.

Applying Logics of Plausible and Implausible Deniability to Influence Operations

Influence Operations and Plausible Deniability

To what extent do the rationales for plausible and implausible deniability from the kinetic covert action literature apply to influence operations? The first hypothesized benefit of Poznansy's "State Model," as shown in Figure 2, is that secrecy decreases the likelihood of retaliation from the target state. For the escalatory rationale for plausible deniability to hold, we would expect two things: First, leaders in a perpetrator country should perceive it as plausible that the target state might retaliate or escalate. If escalation is out of the question, then the logic

[55] Myrick, Rachel. 2020. "Why So Secretive? Unpacking Public Attitudes toward Secrecy and Success in US Foreign Policy." *Journal of Politics* 82(3): 828-843.

does not hold. Second, leaders in a perpetrator country should perceive that the public in the target country would disapprove of interference and support retaliation.

On the one hand, scholars have argued that the probability of influence operations spurring violent retaliation is limited.[56] Violent retaliation to propaganda is empirically rare. However, we have both limited research on whether kinetic retaliation (both violent and non-violent) to propaganda would be supported by the American public, and whether citizens would support non-kinetic retaliation (diplomatic steps, naming and shaming) in response to foreign influence operations.[57]

Forms of retaliation short of significant violence are plausible: in recent years, states have changed their behavior in response to influence operations. For example, U.S. Cyber Command adopted a "persistent engagement" or "defend forward" framework after Russian interference in 2016; in the run-up to the 2018 midterm elections, U.S. Cyber Command blocked Internet access for the Internet Research Agency in St. Petersburg.[58] In April 2021, the Biden Administration unveiled new sanctions on Moscow for its interference in the 2020 presidential election and for a number of cyber-attacks.[59] In both of these cases, response was not in a tit-for-tat form. For the United States, it would not make strategic sense to respond to Russian trolling of U.S. politics with equivalent trolling of Russian politics—both because elections in Russia are less contested and because the United States would face reputational backlash. This dovetails with the legal principle of reprisal, where damage has to be similar but harm does not have to take place in the same domain. Retaliation to influence operations compared to traditional military activity is less likely to come in the same form of the original attack, but may come in the form of diplomatic expulsion or sanctions.

---

[56] Lowenthal, Mark M. 2012. *Intelligence: From Secrets to Policy.* CQ Press; Radsan, A. John. 2009. "An Overt Turn on Covert Action." *Saint Louis University Law Journal.* 53(2): 495.

[57] For an exception, I discuss Tomz and Weeks 2020 further below.

[58] Nakashima, Ellen. "U.S. Cyber Command operation disrupted Internet access of Russian troll factory on day of 2018 midterms." *The Washington* Post. February 27, 2019. Accessed: https://www.washingtonpost.com/world/national-security/us-cyber-command-operation-disrupted-internet-access-of-russian-troll-factory-on-day-of-2018-midterms/2019/02/26/1827fc9e-36d6-11e9-af5b-b51b7ff322e9_story.html; Borghard, Erica. 2019. "What A U.S. Operation Against Russian Trolls Predicts About Escalation in Cyberspace." *War on the Rocks*. Accessed: https://warontherocks.com/2019/03/what-a-u-s-operation-against-russian-trolls-predicts-about-escalation-in-cyberspace/

[59] Smith, Alexander, Kristen Welker, Andrea Mitchell, and Abigail Williams. "Biden calls for de-escalation with Russia, following sanctions, proposes meeting with Putin." *NBC News.* April 15, 2021. Accessed: https://www.nbcnews.com/news/world/u-s-sanction-russia-alleged-election-interference-solarwinds-hack-n1264142

The second criterion—public disapproval of foreign interference or public support for retaliation—is met by research in Political Science. Tomz and Weeks 2020 conducted survey experiments to better understand Americans' perceptions of different forms of foreign electoral interference. In treatment conditions in which a foreign state conducted operations to support its favored candidate—such as spreading truths, spreading lies, giving money, or hacking voting machines—between 72% and 79% of respondents disapproved of the interference. This was significantly higher than rates of disapproval when a foreign government endorsed a candidate (37%) or threatened consequences if its favored candidate did not win (55%). The authors then find that support for various forms of retaliation—including diplomatic sanctions, economic sanctions, military threats, and military strikes—is higher in the operations treatments than other treatments.[60] This provides us with a prima facie expectation that the American public both disapproves when a foreign government wages a specific type of influence operation and supports limited retaliatory measures. Based on this work, we have good reason to believe the retaliatory or escalatory rationale of plausible deniability may hold in the influence operations domain.

> *Hypothesis 1: Citizens in the target state will be less likely to support retaliating against a foreign influence operation when it is not known if a foreign government was involved, compared to when the foreign government operated or paid for the influence operation.*

The second theorized benefit of Poznansky's State Model[61] is that plausible deniability allows a state to avoid the perception of violating international laws and norms. The former is unlikely to motivate plausible deniability in the realm of influence operations, though the latter is applicable.

As Forcese writes, "a state's use of physical force on the territory of another state (such as an arrest and abduction)" violates international law.[62] "Covert action is regulated by international law to the extent it amounts to coercive interference into the affairs of another state

[60] Tomz and Weeks 2020. Note that although support for retaliation was highest in the operation treatment across types of retaliation, there was wide variation in support based on the *type* of retaliation: diplomatic sanctions (59% in operation treatment); economic sanctions (70% in operation treatment); support for military threats (28% in operation treatment) and military strikes (19% in operation treatment).
[61] Poznansky 2020.
[62] Forcese, Craig. 2016. "Pragmatism and Principle: Intelligence Agencies and International Law." *Virginia Law Review Online* 102: 75.

or the nonconsensual exercise of state powers on the territory of another state."[63] Some forms of covert action clearly meet this threshold: if a state assassinates someone or sabotages a powerplant, this is an obvious violation of international law that would incur state responsibility. Influence operations conducted from outside the territory of the target state—particularly those that do not involve cyber intrusions—seem unlikely to meet this threshold.[64] The scholarly consensus is that, at present, it is not universally accepted that states engaging in covert informational campaigns violate international law.[65] As a result, states are unlikely to suffer significant repercussions for their influence operations based on violating international law.[66]

Although violation of international law is unlikely to be a reason to opt for plausible deniability, violation of international norms may be. Covert influence operations violate norms of transparency and noninterference in the domestic affairs of other states. Some government officials throughout the 20th century opposed covert influence operations for this very reason.[67] In addition to hesitancy by government officials, citizens hold an expectation of non-interference. As described above, Tomz and Weeks find that the American public disapproves of even modest form of interference.[68]

> *Hypothesis 2: Citizens in the target state will have a less favorable view of the perpetrator country when the foreign government operated or paid for the influence operation compared to when it is not known if the foreign government was involved.*

[63] Ibid 80.
[64] Schmitt, Michael N. 2017. "Grey Zones in the International Law of Cyberspace." *The Yale Journal of International Law*. 42(2): 1-21.
[65] Goldsmith, Jack and Stuart Russell. 2018. "Strengths Become Vulnerabilities: How a Digital World Disadvantages the United States In Its International Relations." *Hoover Institution Paper*. Aegis Series Paper No. 1806. Accessed: https://www.hoover.org/sites/default/files/research/docs/381100534-strengths-become-vulnerabilities.pdf; Ossoff, William. 2021. "Hacking the Domaine Réservé: The Rule of Non-Intervention and Political Interference in Cyberspace." *Harvard International Law Journal* 62(1): 295-323; Sander, Bernie. 2019. "Democracy Under The Influence: Paradigms of State Responsibility for Cyber Influence Operations on Elections." *Chinese Journal of International Law.* 18(1): 1-56.
[66] While influence operations are unlikely to violate international law, they may violate laws in the target country. For example, the United States has indicted Russians involved in interfering in the 2016 election. Practically, however, this has limited repercussions: Perpetrators won't see their day in American court, barring an extradition treaty and unlikely cooperation at the state level (Sullivan 2016).
[67] https://history.state.gov/historicaldocuments/frus1951-54Iran/d8
[68] Tomz and Weeks 2020.

Influence Operations and Implausible Deniability

In addition to positing benefits of plausible deniability, recent work also suggests that there may be benefits to implausible deniability—when there is substantial evidence linking a perpetrator to an operation, such that a perpetrator's denial is not believable. Cormac and Aldrich write that "ambiguity and implausible deniability allow the construction of powerful narratives."[69] Unlike in some of the other literature on covert action, Cormac and Aldrich apply this to influence operations directly. These logics are summarized in the right side of Figure 2.

Creating an outsized perception of effect can come in both localized and distributed forms. For example, a state waging an influence operation could attempt to impart the perception that the operation is more effective than it truly is. This could undermine confidence in the result of an election, for example. A second form of inflated effectiveness can come in a more distributed form. Instead of inflating perception of a state's impact in a specific operation, it can stir paranoia that the perpetrator is involved in many different efforts or is omnipotent. Facebook terms this "perception hacking," which it defines as when "malicious actors attempt to play on our collective expectation of wide-spread interference to create the perception that they're more impactful than they in fact are."[70]

Both the localized and distributed logics are evident in the most famous recent case of a foreign influence operation: Russian interference in the 2016 election. As of March 2017, 41% of Americans believed that Russia influenced the results of the 2016 election. This is noteworthy given that many of the most rigorous scholarly treatments of the subject find that Russian interference did occur but evidence that it had a significant impact on results is weak.[71] In addition to an inflated localized view of the single operation, Russian interference in 2016 spurred widespread paranoia about Russia's influence over American public opinion writ large.[72] Claire Wardle captures this feeling when she writes: "It's hard to remember when that journalistic skepticism of foreign interference morphed into a zealous belief that Russians were

[69] Cormac and Aldrich 2018.
[70] Gleicher, Nathaniel. "Removing Coordinated Inauthentic Behavior." *Facebook Blog*. October 27, 2020. Accessed: https://about.fb.com/news/2020/10/removing-coordinated-inauthentic-behavior-mexico-iran-myanmar/
[71] Benkler, Yochai, Robert Faris, and Hal Roberts. 2018. *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.
[72] Chen, Adrian. "The Real Paranoia-Inducing Purpose of Russian Hacks." *The New Yorker*. July 27, 2016. Accessed: https://www.newyorker.com/news/news-desk/the-real-paranoia-inducing-purpose-of-russian-hacks

under every bed."[73] One poll in the lead-up to the 2020 election found that 68% of Biden supporters were "extremely" or "very" concerned about foreign countries influencing how Americans perceive the candidates; while 30% of Trump supporters were "extremely" or "very" concerned and 29% "somewhat concerned."[74] States may produce the perception of omnipotence or stir paranoia in the target state when their influence operations are implausibly deniable.

Based on this literature, influence operations that are implausibly deniable may be seen as more effective, they may be more likely to stir paranoia, and the perpetrator may be seen as more powerful and behind a greater number of campaigns compared to when an influence operation is plausibly deniable.

Hypothesis 3: Citizens in the target state will view the perpetrator country as more powerful when the foreign government operated or paid for the operation compared to when it is not known if the foreign government was involved.

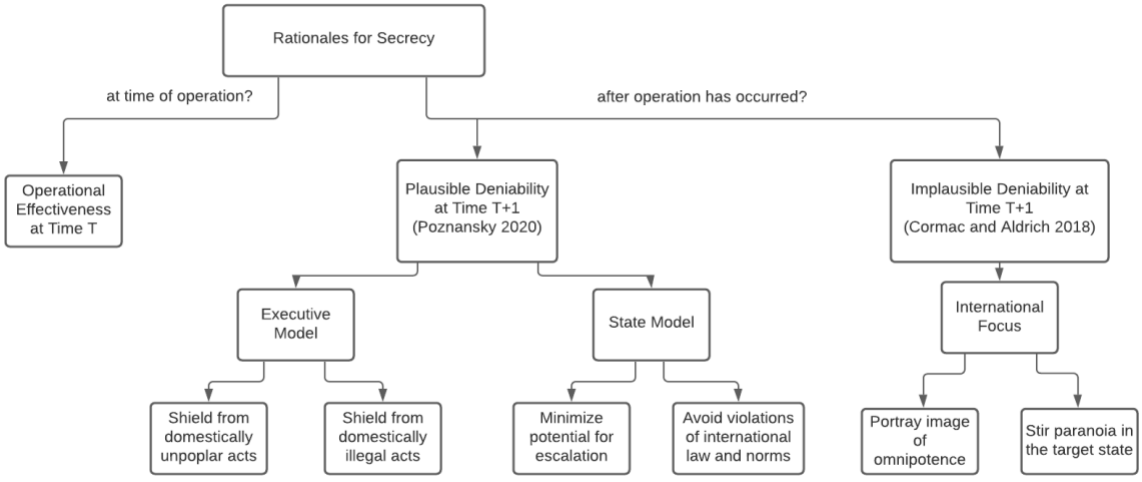
Hypothesis 4: Citizens in the target state will believe an influence operation changed more people's minds and had a bigger effect on the outcome of an election when a foreign government operated or paid for the operation compared to when it is not known if a foreign government was involved.

Hypothesis 5.1: Citizens in the target state will believe that the government in the country of origin of the perpetrator is monitoring their online activities to a greater extent when the foreign government operated or paid for the operation compared to when it is not known if the foreign government was involved.

Hypothesis 5.2: Citizens in the target state will believe that the government in the country of origin of the perpetrator has a greater influence on the day-to-day operations of the U.S. government when the foreign government operated or paid for the operation compared to when it is not known if the foreign government was involved.

[73] Wardle, Clare. 2020. "The Media Has Overcorrected on Foreign Influence." *Lawfare.* Accessed: https://www.lawfareblog.com/media-has-overcorrected-foreign-influence
[74] Associated Press. "Three in four Americans are worried about foreign interference in the presidential election." *The Associated Press*. October 2, 2020. Accessed: https://www.marketwatch.com/story/most-americans-are-worried-about-foreign-interference-in-the-election-01601648756

Figure 2: Summary of Rationales for Secrecy in the Covert Action Literature



Secondary Expectations: The Effects of Partisanship and Country Favorability

In addition to testing the logics of plausible and implausible deniability discussed above, I also develop secondary expectations about the possible effects of prior views of the country of origin and partisanship.

We do not expect all citizens to react to allegations of foreign influence operations under conditions of uncertainty in the same way. Assignment of blame and support for retaliation is likely to be influenced by people's prior views of the country of origin of a perpetrator. Research in psychology suggests that how people view the actions and proposals of others is based on their prior views of them. For example, research on peace proposals suggests that citizens view an identical proposal as more favorable when it is purportedly proposed or supported by their in-group members, and significantly less favorable when it is proposed or supported by out-group members.[75] If people assign blame and make judgements based on their prior views, we should expect respondents to be more likely to assign blame to the foreign government, even if a foreign government's role is uncertain, when they hold an unfavorable view of the foreign country.

[75] Maoz, Ifat, Andrew Ward, Michael Katz, and Lee Ross. 2002. "Reactive Devaluation of an 'Israeli' vs. 'Palestinian' Peace Proposal." *Journal of Conflict Resolution.* 46(4): 515-546. For a study on how the identity of an actor interfering in elections impacts public perceptions of the intervention, see: Shulman, Stephen and Stephen Bloom. 2012. "The legitimacy of foreign intervention in elections: the Ukrainian response." *Review of International Studies*. 38(2): 445-471.

Hypothesis 6: Citizens in the target state will be more likely to believe that a foreign government played a role in a foreign influence operation if the country of origin of the operation has low favorability in the United States (e.g. Russia or China) compared to if it has high favorability (e.g. the United Kingdom or Japan).

Hypothesis 7: Citizens in the target state will be more likely to support retaliation if the country of origin of a foreign influence operation has low favorability in the United States (e.g. Russia or China) compared to if it has high favorability (e.g. the United Kingdom or Japan).

Hypothesis 8: The effect of plausible versus implausible deniability on the belief that a foreign government was involved in a foreign influence operation stemming from their country will be smaller when the country of origin has low favorability in the United States (e.g. Russia or China) compared to when it has high favorability (e.g. the United Kingdom or Japan).[76]

Citizens in the target state may also react differently to an influence operation based on whether it helps or hurts their favored candidate. Corstange and Marinov used survey experiments in the wake of Lebanon's 2009 parliamentary elections to study public perceptions of foreign interference by the United States and Iran. The authors found strong partisan alignment: Sunnis supportive of the Western-backed alliance responded favorably to American interference, while Shia supporters of the Iranian-backed bloc responded unfavorably.[77] This study considered foreign electoral interference, but not covert efforts. In a more recent study, Tomz and Weeks found that Americans responded much more negatively to foreign electoral interference when the intervention supported the opposition than when it supported the respondent's favored party.[78]

Literature in political psychology on partisan motivated reasoning suggests that partisans process new information through the lens of their party commitments.[79] Under conditions of uncertainty, we would expect people to be more likely to blame a foreign power for the operation when it hurts their favored candidate compared to when it hurts the opposition. In other words,

---

[76] I operationalize plausible and implausible deniability as when a foreign government operates or pays for an influence operation (implausibly deniable) compared to when it is not known if the foreign government was involved (plausibly deniable).
[77] Corstange, Daniel and Nikolay Marinov. 2012. "Taking Sides in Other People's Elections: The Polarizing Effect of Foreign Intervention." *American Journal of Political Science* 56(3): 664.
[78] Tomz and Weeks 2020, p. 8.
[79] Bolsen, Toby, James N. Druckman, and Fay Lomax Cook. 2014. "The Influence of Partisan Motivated Reasoning on Public Opinion." *Political Behavior* 36(2): 235-262; Bisgaard, Martin. 2019. "How Getting the Facts Right Can Fuel Partisan-Motivated Reasoning." *American Journal of Political Science* 63(4): 824-839.

uncertainty will 'matter less' when assigning blame when a partisan's own team is hurt than when the opposition is hurt. Second, we would expect citizens to exhibit higher levels of support for retaliation to operations that hurt their candidate compared to operations that hurt the opposition; this should hold both under conditions of plausible deniability and implausible deniability.

> *Hypothesis 9: Citizens in the target state will be more likely to believe that a foreign government played a role in a foreign influence operation if the operation spreads content that favors an opposition candidate compared to if it spreads content that favors their preferred candidate.*
>
> *Hypothesis 10: Citizens in the target state will be more likely to support retaliation if a foreign influence operation spreads content that favors an opposition candidate compared to if it spreads content that favors their preferred candidate.*
>
> *Hypothesis 11: The effect of plausible versus implausible deniability on the belief that a foreign government was involved in a foreign influence operation stemming from their country will be smaller when an operation criticizes the respondent's favored candidate than when an operation criticizes the opposition.*[80]

Experimental Design

I designed a survey experiment (n=7,792) that I deployed using Lucid in June-July 2021.[81] Lucid Theorem is a panel provider that creates samples designed to approximate the U.S. population.[82] Many recent papers in Political Science and misinformation studies use Lucid to construct samples.[83] Coppock and McClellan find very similar treatment effect estimates when they replicated on Lucid existing studies; the authors conclude that Lucid is an appropriate

[80] I operationalize plausible and implausible deniability as when a foreign government operates or pays for an influence operation (implausibly deniable) compared to when it is not known if the foreign government was involved (plausibly deniable).
[81] For research showing that Lucid provides a sample suitable for testing social science theories, see. For examples of papers using Lucid, see Coppock, Alexander and Oliver A. McClellan. 2019. "Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents." *Research and Politics* 1-14; Orr, Lilla V. and Gregory A. Huber. 2020. "The Policy Basis of Measured Partisan Animosity in the United States." *American Journal of Political Science* 64(3): 569-586; Tomz and Weeks 2020.
[82] They quote sample based on age, gender, ethnicity and region.
[83] For a few examples, see: Bode, Leticia and Emily K. Vraga. 2021. "Correction Experiences on Social Media During COVID-19." *Social Media + Society*: 1-11. Robinson, Joshua. 2021. "Can Elites Escape Blame by Explaining Themselves? Suspicion and the Limited of Elite Explanations." *British Journal of Political Science*: 1-20. Stecula, Dominik A. and Mark Pickup. 2021. "How populism and conservative media fuel conspiracy beliefs about COVID-19 and what it means for COVID-19 behaviors." *Research & Politics*: 1-9.

sample provider for studies estimating sample average treatment effects (SATEs).[84] All expectations and the empirical strategies in the results section below were pre-registered with Evidence in Governance and Politics (EGAP) prior to the deployment of the survey.[85]

The survey was run on 7,792 respondents. Each of the treatments was assigned equal numbers of respondents. The sample was 46.9% male and 77.2% White (12.4% Black or African American, 11.5% Hispanic, and 4.8% Asian). The mean respondent age was 46.3 years, and 36.1% of the sample completed a 4-year or postgraduate degree; 28.6% of respondents think of themselves as Republicans and 36.5% as Democrats. For more details on demographic questions, see the survey questionnaire included in the Appendix.

In my survey, I ask respondents about a future scenario that could take place and vary attribution of a foreign social media influence operations targeting the 2024 U.S. election. This helps unpack public responses to influence operations targeting elections under different plausible conditions.[86]

To create plausible deniability, I vary whether a foreign government operated and/or paid for a foreign influence operation. These are distinct dimensions, given that one can outsource disinformation work to another entity to operate the accounts. For who conducted the operation, I vary whether the U.S. intelligence community assesses that it was operated by a foreign government, a foreign marketing firm, or it is unknown who operated it. As for who sponsored the operation, I vary whether the U.S. intelligence community assesses that it was paid for by the foreign government, private actors, or unknown sources. I exclude unrealistic combinations, such as private citizens paying the foreign government to conduct an operation. For an exposition of the attribution treatments, see Table 1.

My theoretical expectations derived above offer expectations for how Americans will view operations that are implausibly deniable and those that are plausibly deniable. Of note, my theory did not produce expectations that citizens in the target state will respond differently based on whether an operation is run by a government or paid for by a government (implausible deniability treatments) or based on the treatments that all produce plausible deniability.

[84] Coppock and McClellan 2019.
[85] See: https://osf.io/7ufxy
[86] Future research could look at how conditions of uncertainty impact public perceptions of foreign influence operations that relate to elections and that do not relate to elections.

Differences among plausible deniability treatments and among implausible deniability treatments are interesting for policy,[87] but my theory only leads to one axis of difference.

Table 1: Treatments for Who Conducted and Who Directed the Foreign Influence Operation

The U.S. intelligence community…

		Who Carried Out the Operation?		
		Government	*Marketing Firm*	*Uncertain*
Who Directed the Operation?	*Government*	assessed that the [*country*] government operated the accounts. *(Implausible Deniability)*	assessed that a marketing firm in [*country*] operated the accounts and that the marketing firm was paid by the [*country*] government to do so. *(Implausible Deniability)*	*Implausible Scenario, Not Treated*
	Private Citizens	*Implausible Scenario, Not Treated*	assessed that a marketing firm in [*country*] operated the accounts and that the marketing firm was paid by private [*country*] citizens to do so. *(Plausible Deniability)*	*Implausible Scenario, Not Treated*
	Uncertain	*Implausible Scenario, Not Treated*	assessed that a marketing firm in [*country*] operated the accounts, but the U.S. intelligence community does not know who paid the marketing firm to do so. *(Plausible Deniability)*	does not know who operated the accounts or paid for the operation. *(Plausible Deniability)*

In addition to varying the attribution, I vary two additional components of the vignette. First, I vary whether the operation supported the Democratic candidate and criticized the Republican candidate, or whether it supported the Republican candidate and criticized the Democratic candidate. This allows me to test whether respondents exhibit partisan double-standards for blame and retaliation.

Second, I vary whether the foreign country in question is China, Russia, the United Kingdom, or Japan. I select Russia and China because they are realistic cases. Based on public attributions, Russia is a frequent state source of social media influence operations (including targeting the United States).[88] Chinese influence operations have historically focused on self-promotion of China, not subverting adversaries. However, the Chinese Communist Party has

[87] For example, it is interesting whether Americans perceive operations that are run by the government themselves as more effective than operations that the government pays a marketing firm to run. If one route gives the target an impression that an operation is more effective and the perpetrator's goal is subversion in the target state, then policymakers in the target state should be particularly sensitive to messaging about operations from that source.
[88] e.g. Gleicher et al. 2021.

recently used fake social media accounts to push false messages about COVID-19 to international audiences. Given the CCP's vast resources and growing great-power competition, it is important to understand how Americans will respond to allegations of Chinese influence operations.

Moreover, I include the United Kingdom and Japan in order to create variation in whether countries are viewed favorably or unfavorably by American citizens. China and Russia are high-capability actors that are viewed unfavorably by the United States. The United Kingdom and Japan are high capability countries with high favorability ratings in the United States.[89] According to 2021 YouGov data, 64% of Americans view the United Kingdom favorably and 60% view Japan favorably.[90] By contrast, only 24% view Russia favorably and 23% view China favorably. While it is impossible to provide real-world countries that are identical to China and Russia on all dimensions other than favorability, the United Kingdom and Japan are among the world's top 10 military spenders. The United Kingdom is a nuclear power. Differences in assignment of blame and support for retaliation are thus unlikely to be directed by military capacity alone.

Table 2: Summary of Experimental Treatments

Attributed to (country)	Attributed to (actor)	Operation supports (candidate)
• Russia • China • United Kingdom • Japan	• Government • Marketing firm, paid by government • Marketing firm, paid by private citizens • Marketing firm, don't know who paid them • Control: Don't know who carried it out or paid people to do so	• Democratic candidate • Republican candidate

Vignette

In my vignette, I build on Tomz and Weeks 2020, which presents respondents with a scenario about foreign interference in the 2024 U.S. election.[91] I selected this vignette as a model

[89] See: https://today.yougov.com/ratings/travel/popularity/countries/all
[90] These numbers are high, as, for reference, 77% of Americans view the United States of America positively.
[91] Experiments in International Relations frequently rely on either vignettes or mock news clips. Kreps and Roblin find that "findings across formats may be comparable for all practical purposes." Kreps, Sarah and Stephen Roblin. 2019. "Treatment format and external validity in international relations experiments." *International Interactions* 45(3): 591.

because although foreign influence operations are not exclusive to elections, electoral interference is an important use case.

I first told participants that I would describe a situation that could take place in the future, and then ask them for their opinions. I then presented respondents with a scenario of a foreign influence operation targeting the U.S. election. To study the effect of plausible deniability and outsourcing, I experimentally varied who operated and/or paid for the operation.

> In 2024, during the U.S. Presidential election campaign, a network of thousands of fake accounts on social media pretended to be real Americans and shared content that favored the [*candidate*] and criticized the [*opposition*].
>
> The U.S. intelligence community:
> - assessed that the [*country*] government operated the accounts.
> - assessed that a marketing firm in [*country*] operated the accounts and that the marketing firm was paid by the [*country*] government to do so.
> - assessed that a marketing firm in [*country*] operated the accounts, but the U.S. intelligence community does not know who paid the marketing firm to do so.
> - assessed that a marketing firm in [*country*] operated the accounts and that the marketing firm was paid by private [*country*] citizens to do so.
> - does not know who operated the accounts or paid for the operation.
>
> A spokesperson from the [*country*] government has stated that the [*country*] government was not involved in the operation.[92]

In the scenario, *candidate* and *opposition* were randomly assigned to be "Democratic candidate" and "Republican candidate" or vice versa. *Country* was randomly assigned to be Russia, China, the United Kingdom, or Japan. I conjugated country appropriately (e.g. Russia vs. Russian). Demographic questions were collected prior to the vignette. After the vignette, I measured opinion in line with the developed hypotheses. First, would the actor behind the operation change respondents' support for retaliation? (escalation logic) Second, would the actor behind the operation change respondents' views of the country from which the operation originated? (norms logic) Third, would the actor behind the operation change the respondents' view of the success of the operation and the power of the country from which the operation

[92] The control group (the treatment that is told: "The U.S. intelligence community does not know who operated the accounts or paid for the operation") does not receive the final sentence in the vignette ("A spokesperson from the [*country*] government has stated that the [*country*] government was not involved in the operation). Logically, the statement would not make sense if no one on the country's soil has been accused of running the operation.

originated? (implausible deniability logics) I include the full survey questionnaire in the Appendix.

Results

Testing the Benefits of Plausible Deniability

I first test two theorized benefits of plausible deniability: that operations that are plausibly deniable will be less likely to engender support for retaliation (H1) and less likely to produce reputational backlash (H2).

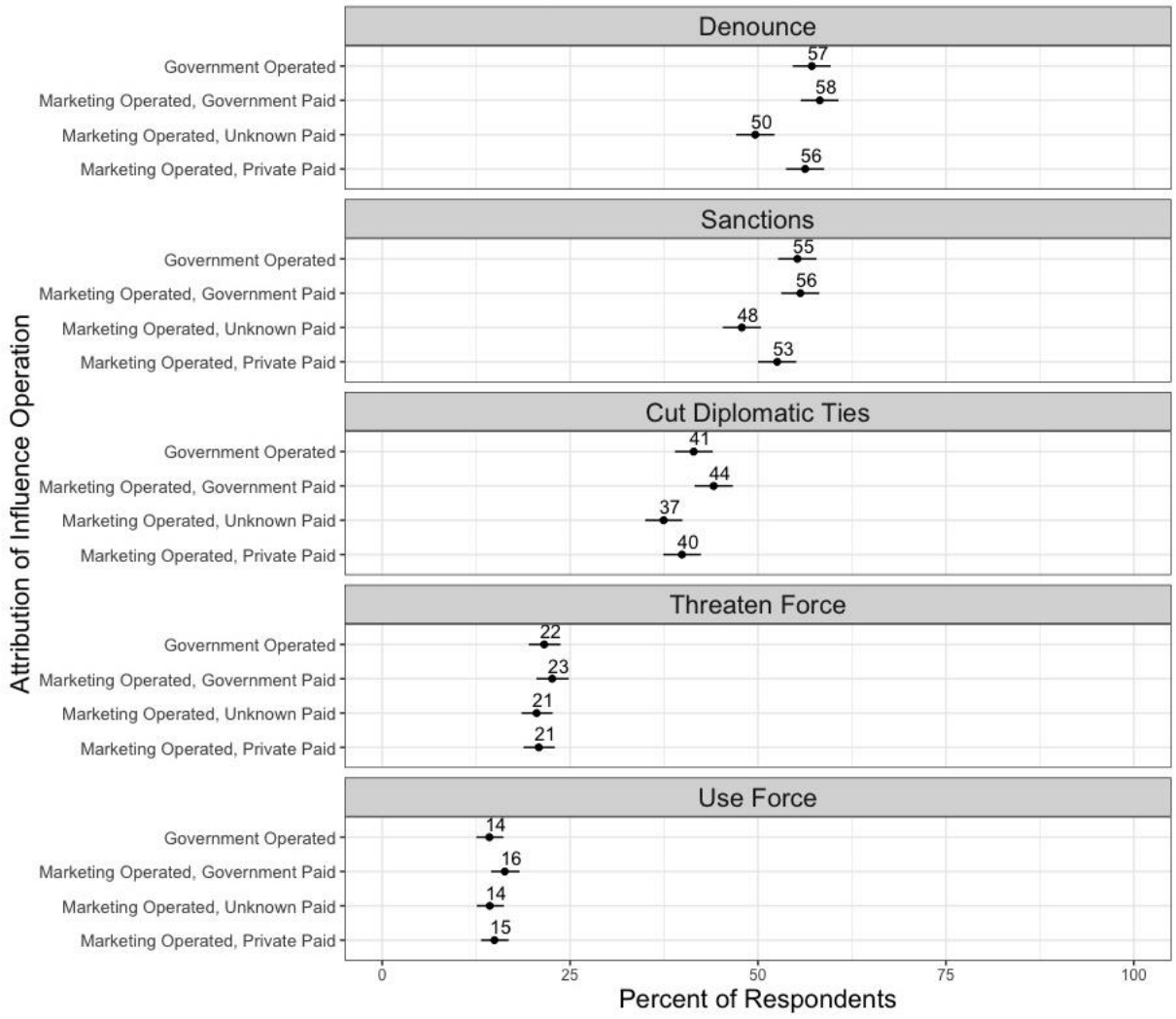
To test Hypothesis 1, I asked respondents whether they would support or oppose different retaliatory measures by the United States government. For each retaliatory measure, I presented respondents with a 4-point scale (from "Strongly oppose" to "Strongly support").[93] Figure 3 shows the percentage of respondents in each treatment group who "Support" or "Strongly support" the proposed retaliatory measure. While there are moderate levels of support for denouncing the country, enacting sanctions, and cutting diplomatic ties, support for threatening force or using force is quite low.

As shown in Figure 3, respondents who are told that the government operated the accounts are significantly more likely to support denouncing (a difference of 7 percentage points), sanctioning (a difference of 7 percentage points), or cutting diplomatic ties with the perpetrator country (a difference of 4 percentage points) compared to the treatment in which a marketing firm operated the accounts but it is unknown who paid for the operation, providing support for H1. The difference between the treatments where the government was involved directly and the treatment where the marketing firm operated the accounts and was paid to do so by private citizens was smaller, perhaps suggesting that people assume that the foreign citizens were government-linked. However, against my expectations, I do not find differences between the treatment groups in terms of their support for threatening to use force or using force. This may be due to a floor effect, as only a small portion of respondents supported drastic escalatory measures even in the group that learned that the foreign government operated the accounts directly. Perpetrators are likely to know that the upper bound of retaliation to social media

[93] I did not ask the control group—the group that was told that the U.S. intelligence community "does not know who operated the accounts or paid for the operation"—questions about retaliating. It would not make sense to ask this group about retaliation against a country if no one from the country was accused of being involved.

influence operations is thus limited, which may help to explain the frequency of such operations.[94]

Figure 3: Percent of Respondents Who Support Retaliation

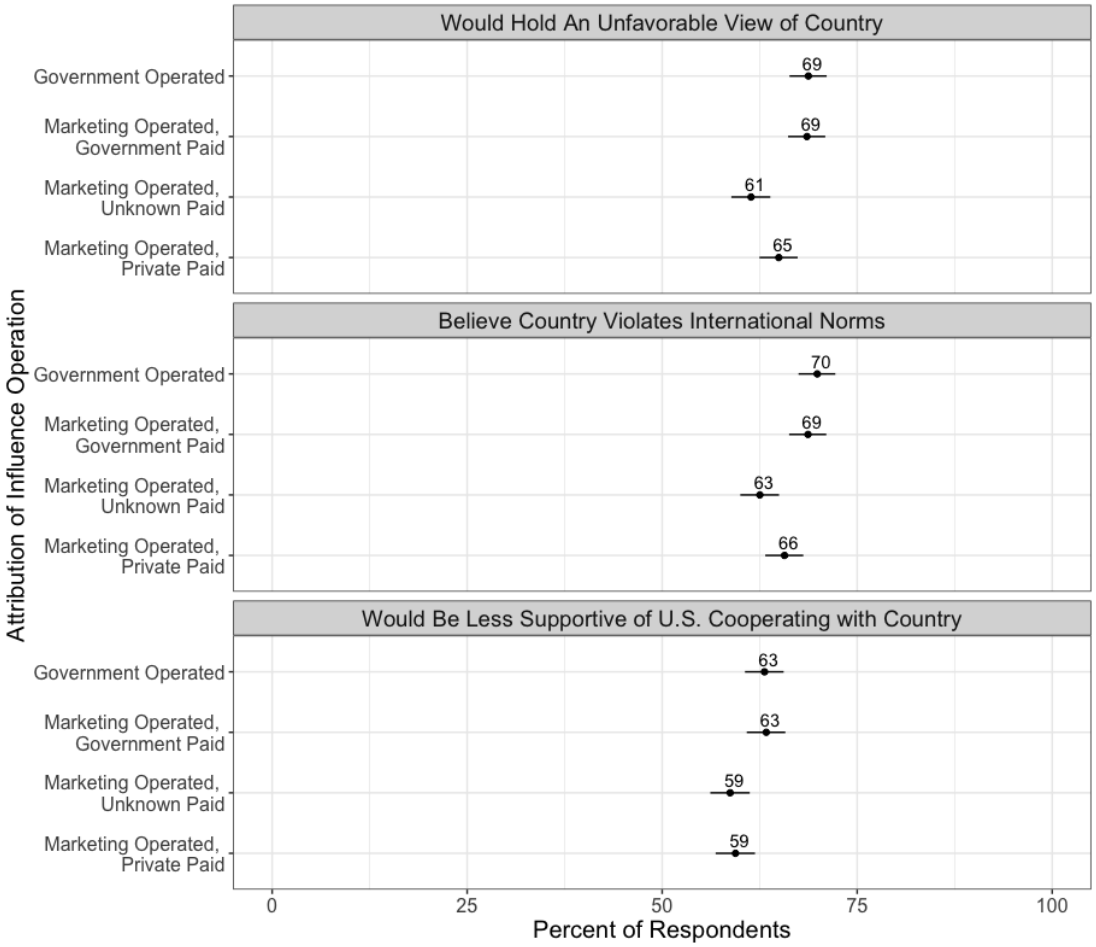


The second theorized benefit of plausible deniability is that citizens in the target state will view a perpetrator country less favorably when the government is involved in an influence operation originating in the country compared to when it is not known whether the government was involved. In Figure 4, I find significant effects between the plausibly deniable and

[94] In the main text, I primarily present results by each treatment group. In the Appendix, I present the same tests on binary treatments (e.g. plausible vs. implausible treatments, ally vs. enemy treatments) given that my theory only predicts one axis of difference for each test. All results between the main text and Appendix are consistent.

implausibly deniable treatment groups across measures. When the government operated the accounts, 69% of respondents said they held an unfavorable view of the country, 70% believed the country violates international norms, and 63% were less supportive of the United States cooperating with the country in the future. When the marketing firm operated the accounts and it is unknown who paid for them, 61% held an unfavorable view of the country, 63% believed the country violates international norms, and 59% were less supportive of the United States cooperating with the country in the future.

Figure 4: Percent of Respondents Who…



Testing the Benefits of Implausible Deniability

Findings from the prior section suggest that states that wage plausibly deniable influence operations benefit both in terms of lower support for retaliation in the target state and weaker reputational costs. Do they forego certain benefits of implausible deniability? In the theory

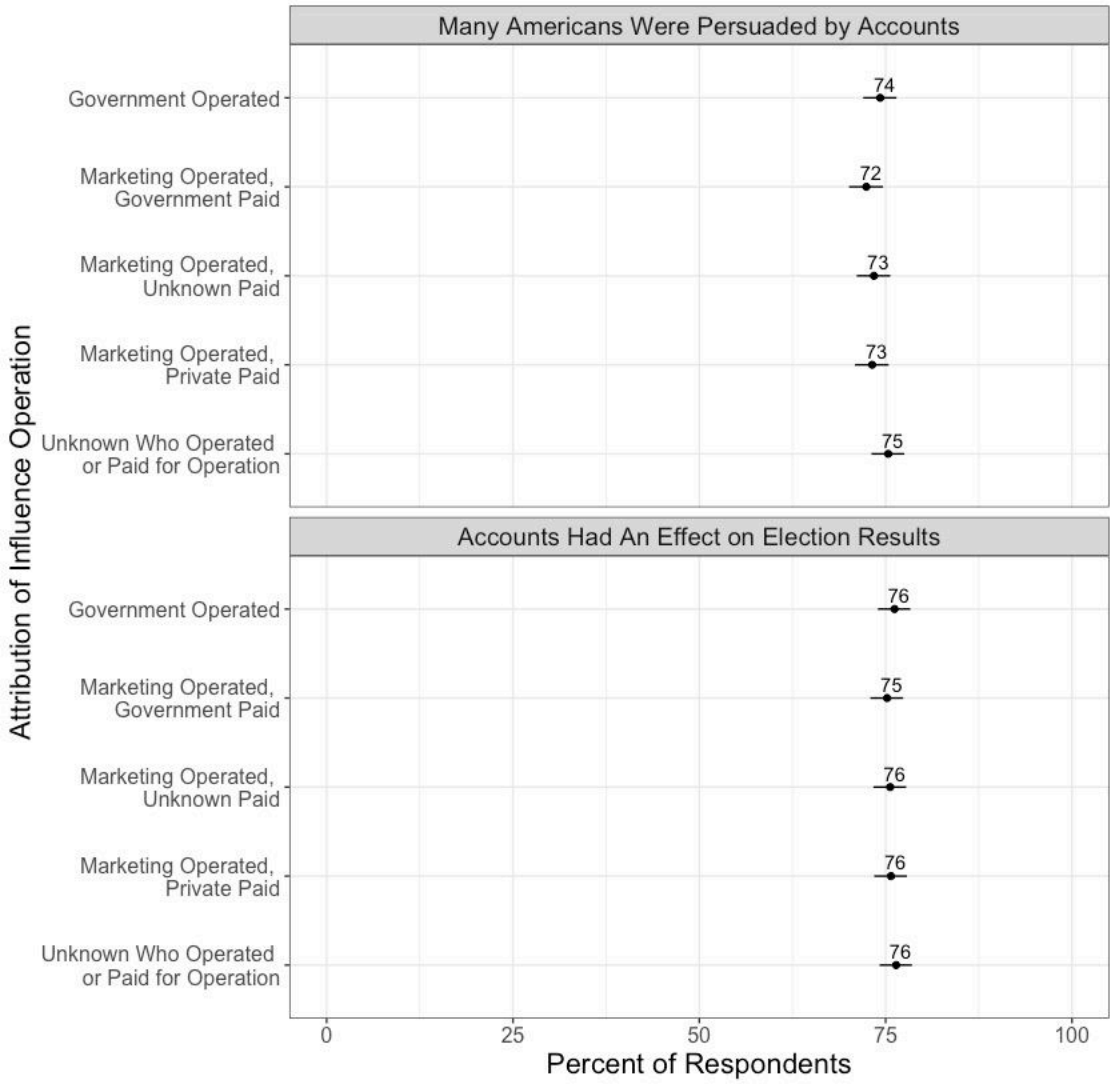
section, I outlined that potential benefits of implausible deniability include that a country is seen as more powerful, that the operation is seen as more effective, and that the operation stirs paranoia about influence operations in the target state.

To test these hypotheses, I asked respondents "If 2024 happened just as we described, to what extent would you agree or disagree with the following statements?" The first set of statements related to the success of the operation:

- Many Americans were persuaded by information coming from these fake social media accounts.
- These fake social media accounts likely had an effect on the election results.

As shown in Figure 5, respondents were no more likely to believe that an operation persuaded Americans or had an effect on the election results based on who operated the accounts. Between 72% and 75% "Agreed" or "Strongly agreed" that many Americans were persuaded by the accounts, with no clear division between plausibly deniable and implausibly deniable treatments.

Figure 5: Percent of Respondent Who Believe That...



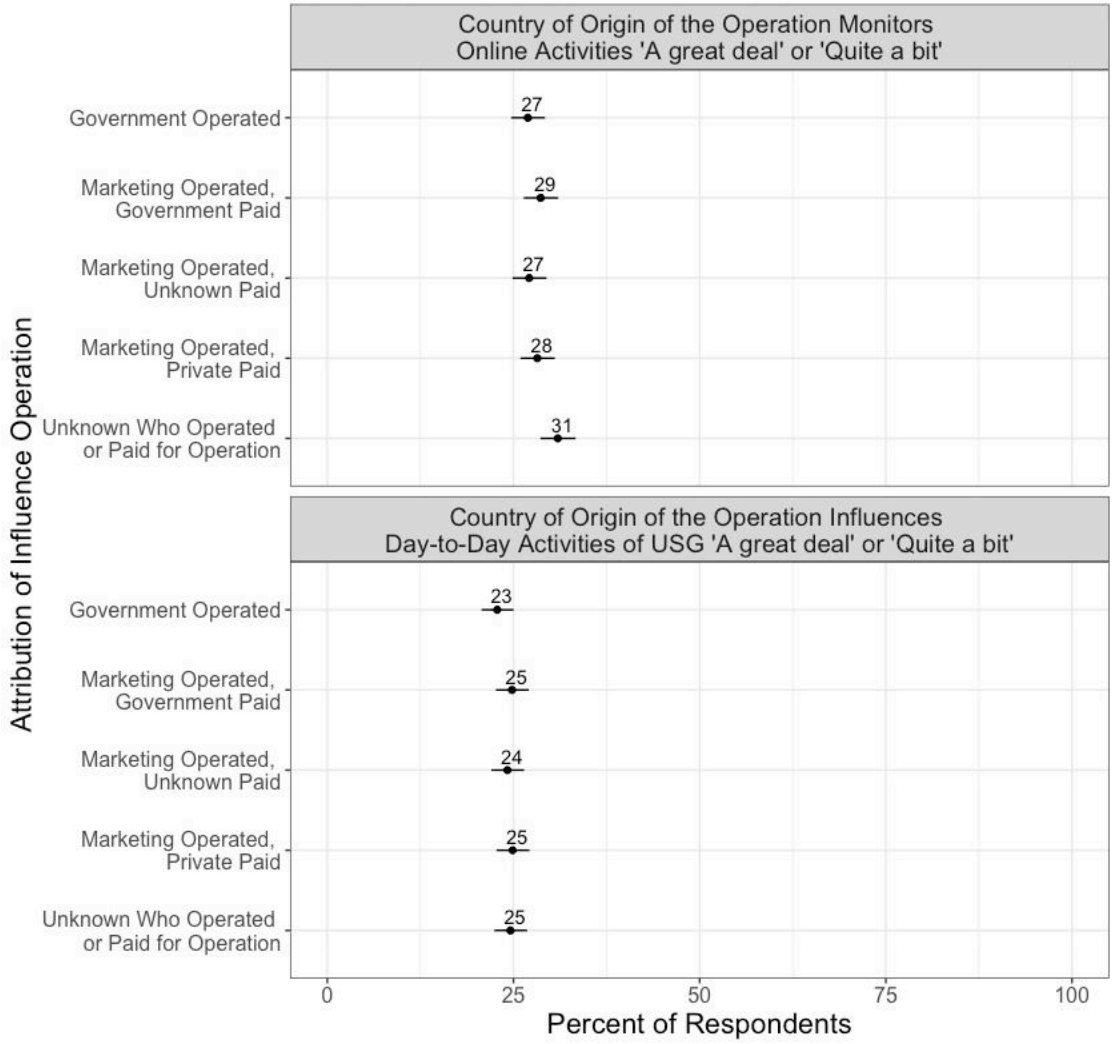
To test whether respondents would be more paranoid of foreign influence operations from the country of origin, I asked:

- To what extent do you think the [*country*] government is monitoring your online activities?
- To what extent does the [*country*] government influence the day-to-day operations of the United States government?

Both questions included 5-point response scales (ranging from "Not at all" to "A great deal"). In Figure 6, I plot the percent of respondents answered a "A great deal" or "Quite a bit." Between 27% and 31% of respondents responded that the foreign government monitors their online activity "A great deal" or "Quite a bit" with no clear difference between plausibly

deniable and implausibly deniable treatments. The percent of respondents who said the foreign government influences the day-to-day activities of the U.S. government "A great deal" or "Quite a bit" varied between 23% and 25%, again, with no clear difference between plausibly deniable and implausibly deniable treatments.

Figure 6: Percent of Respondents Who Believe That...



The primary analysis above focuses on the effects of attribution on the supposed benefits of plausible and implausible deniability. I find significant but moderate benefits of plausible deniability, and no benefits of implausible deniability. These findings help to explain the rising trend of outsourcing influence operations: When government involvement cannot be proven, the country of origin is less likely to suffer reputational damage and citizens in the target state are less likely to support retaliation. At the same time, the operation is no more likely to make the

country seem strong, to stir paranoia, or to give the impression that the operation was effective. As I show in the Appendix, these null findings are consistent when I only include survey respondents who read the scenario closely (operationalized as those who passed post-treatment attention checks.)

Effect of Identity of the Perpetrator

I also investigated how the identity of the country of origin impacts assignment of blame and support for retaliation. To assess blame, I asked respondents in all treatments except for the control group the following two questions:[95]

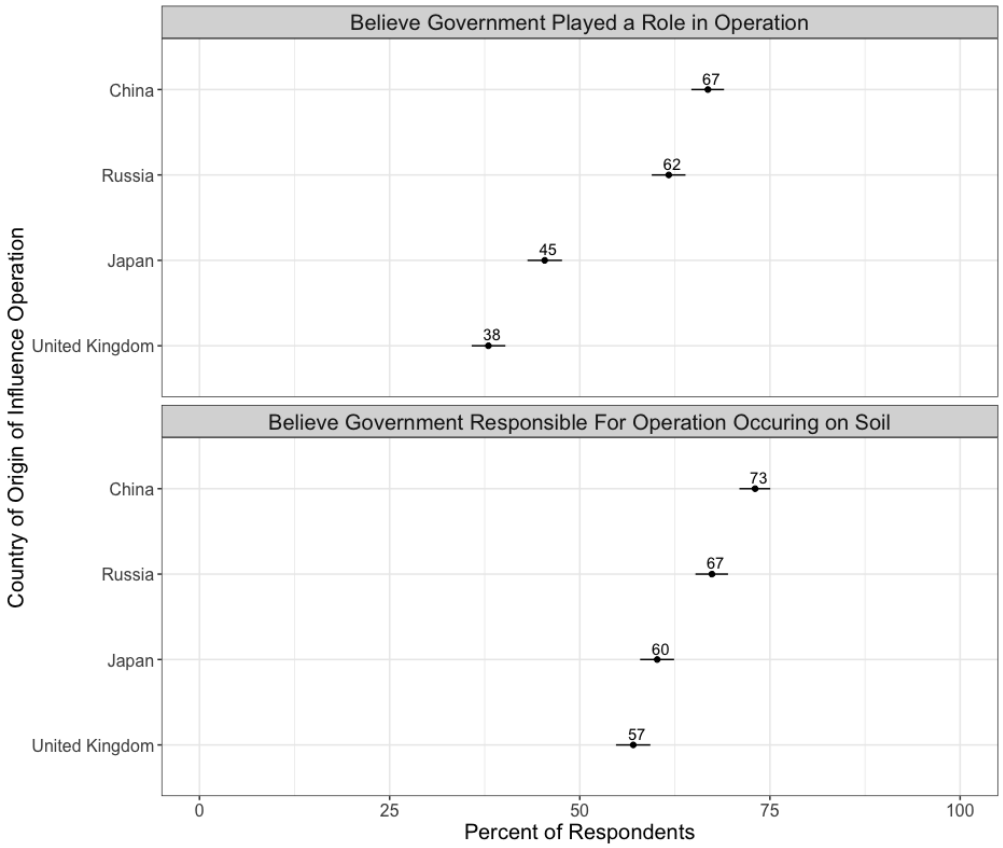
- Do you believe that the [*country*] government played a role in the operation?
- To what extent is the [*country*] government responsible for the operation occurring on [*country*] soil?

Each of these questions were accompanied by 4-point response scales. As shown in Figure 7, Americans are far more likely to believe that a country's government played a role in an operation when the country of origin of the operation is an adversary compared to an ally. When China or Russia was the country of origin, 67% and 62% of respondents, respectively, believed that the government played a role in the operation; when Japan or the United Kingdom was the country of origin, 45% and 38% of respondents believed the government played a role.

When asked about whether respondents believed the country was responsible for the operation occurring on its soil, the identity of the country had a smaller, but still substantive effect: 73% believed the government was responsible when the operation originated in China, 67% when it originated in Russia, compared to 60% when it originated in Japan and 57% when it originated in the United Kingdom.

[95] I did not ask the control group—the group that was told that the U.S. intelligence community "does not know who operated the accounts or paid for the operation"—questions about a foreign government's involvement or responsibility. It would not make sense to ask this group these questions if no one from the foreign country was accused of being involved.

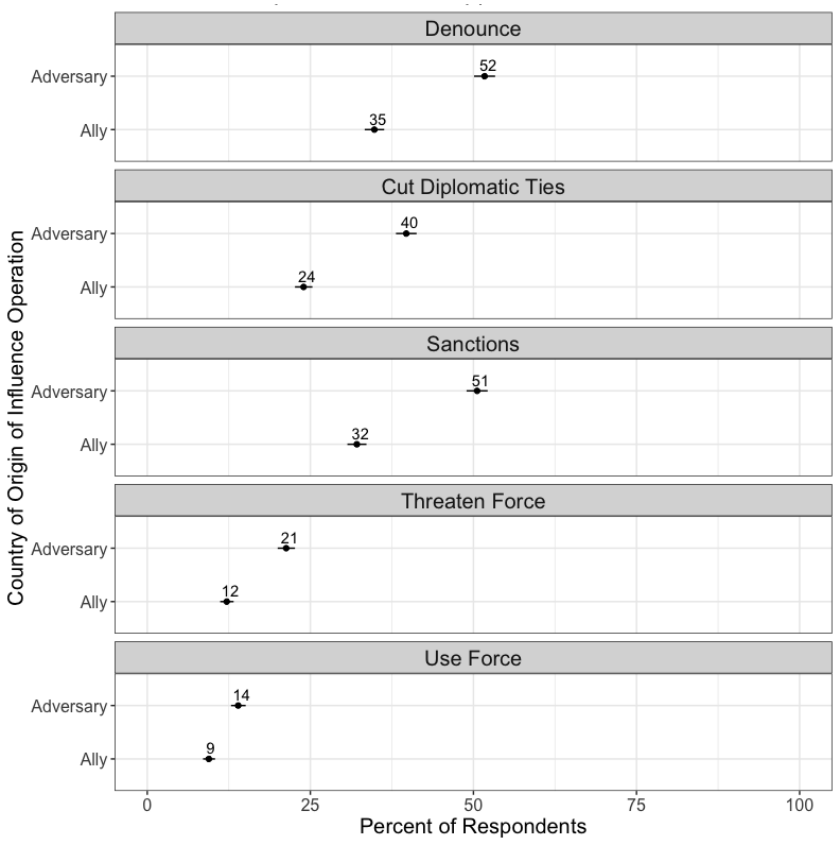
Figure 7: Percent of Respondents Who Believe Foreign Government Was Involved or is Responsible for the Operation



In addition to assignment of blame, I tested whether support for retaliation is conditional on the identity of the perpetrator. Here, I used the same dependent variable questions as in the primary analysis above. For ease of presentation, I grouped favorable countries (Japan and the United Kingdom) and unfavorable ones (China and Russia) together.[96] As shown in Figure 8, the difference in support for retaliation is significant across both lower levels and more drastic forms of retaliation. For example, 51% of respondents who were told that the operation originated in Russia or China supported sanctions compared to 32% when the operation originated in Japan or the United Kingdom, a difference of 17 percentage points.

[96] In the Appendix, I reproduce the results without grouping. The findings are consistent: a higher percentage of respondents are supportive of retaliating when the operation originates in Russia or in China compared to when it originates in Japan or the United Kingdom. Support for retaliation is highest for China, then Russia, followed by Japan and the United Kingdom.

Figure 8: Percent of Respondents Who Support Retaliation



Finally, I investigate whether the effects of plausible deniability on assignment of blame are different for favorable allies (Japan and the United Kingdom) and unfavorable adversaries (China and Russia). As shown in Figure 9, when an operation stems from an adversary, there is no effect of plausible deniability on assignment of blame: 80% believe a government played a role when it is plausibly deniable and 79% believe a government played a role when it is implausibly deniable. Likewise, 87% of respondents believe the government was responsible for the operation occurring on its soil—both when the government was involved and when the government's role was not known. By contrast, when an operation originates in an ally's territory, the difference in percent of respondents who believe the government played a role is 3 percentage points when it is implausibly deniable, and the difference in percent of respondents who believe the government is responsible for the operation occurring on its soil is 6 percentage points. Allies benefit from plausible deniability (albeit with a small effect) while adversaries do not.

Figure 9: Percent of Respondents Who Support Retaliation

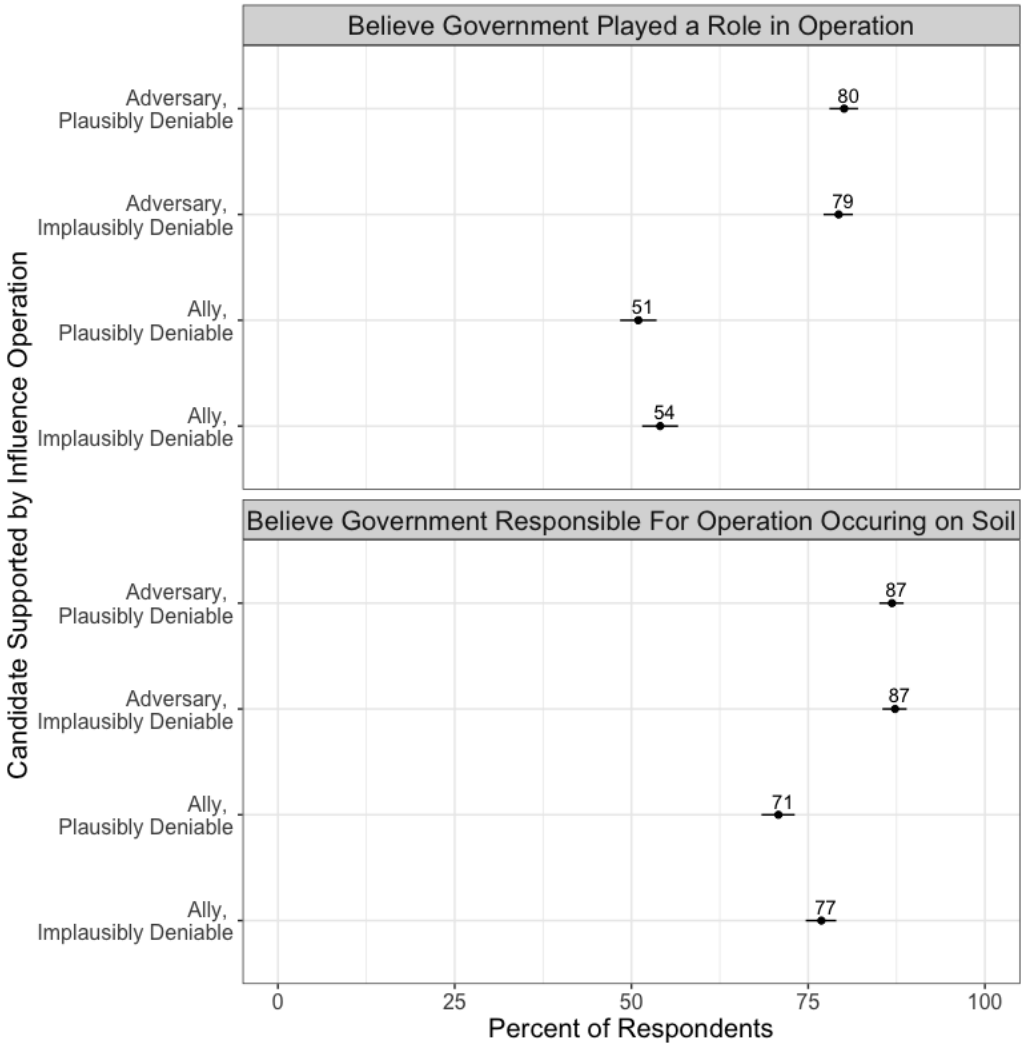
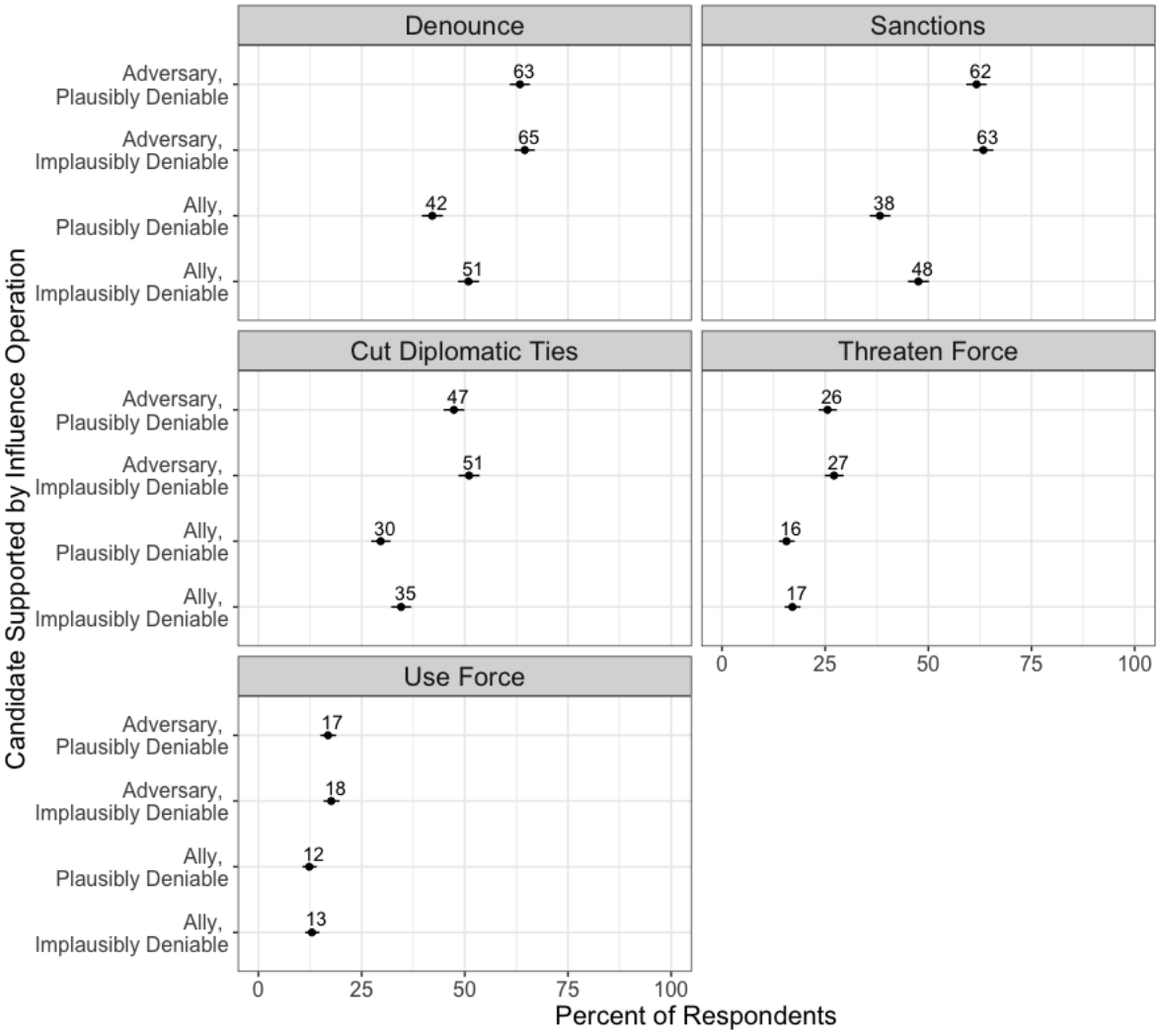


Figure 9 clearly demonstrates that the effect of the country is much larger than the effect of plausible versus implausible deniability. For an adversary, the difference in belief that a government played a role in an operation is 1 percentage point between plausible deniability and implausible deniability treatments. However, the difference in belief that a government played a role in an operation is 29 percentage points between a plausibly deniable operation originating in an adversary country compared to an allied country (80% believe an adversary played a role in an operation even when they are not told that the government played a role, compared to 51% when the same type of operation originates in an allied country). These novel results suggest that it is not so much the way in which countries wage influence operations, but *which* countries wage the influence operation, that determines public reactions in the target state.

I also conduct exploratory analysis on whether there are heterogenous treatment effects on support for retaliation. For lower level forms of retaliation (e.g. denounce, sanctions) respondents are significantly more likely to support the retaliatory measure when an operation is implausibly deniable than when it is plausibly deniable if it originates in an allied country. When it originates in an adversary country, there is not a significant difference in percent of respondents who support retaliating based on plausible deniability.

Figure 10: Percent of Respondents Who Support Retaliation



These findings may help to explain the null finding from the main analysis that plausible deniability does not have a significant effect on whether the country of origin of an operation monitors the respondent's online activities and whether the country of origin influences day-to-

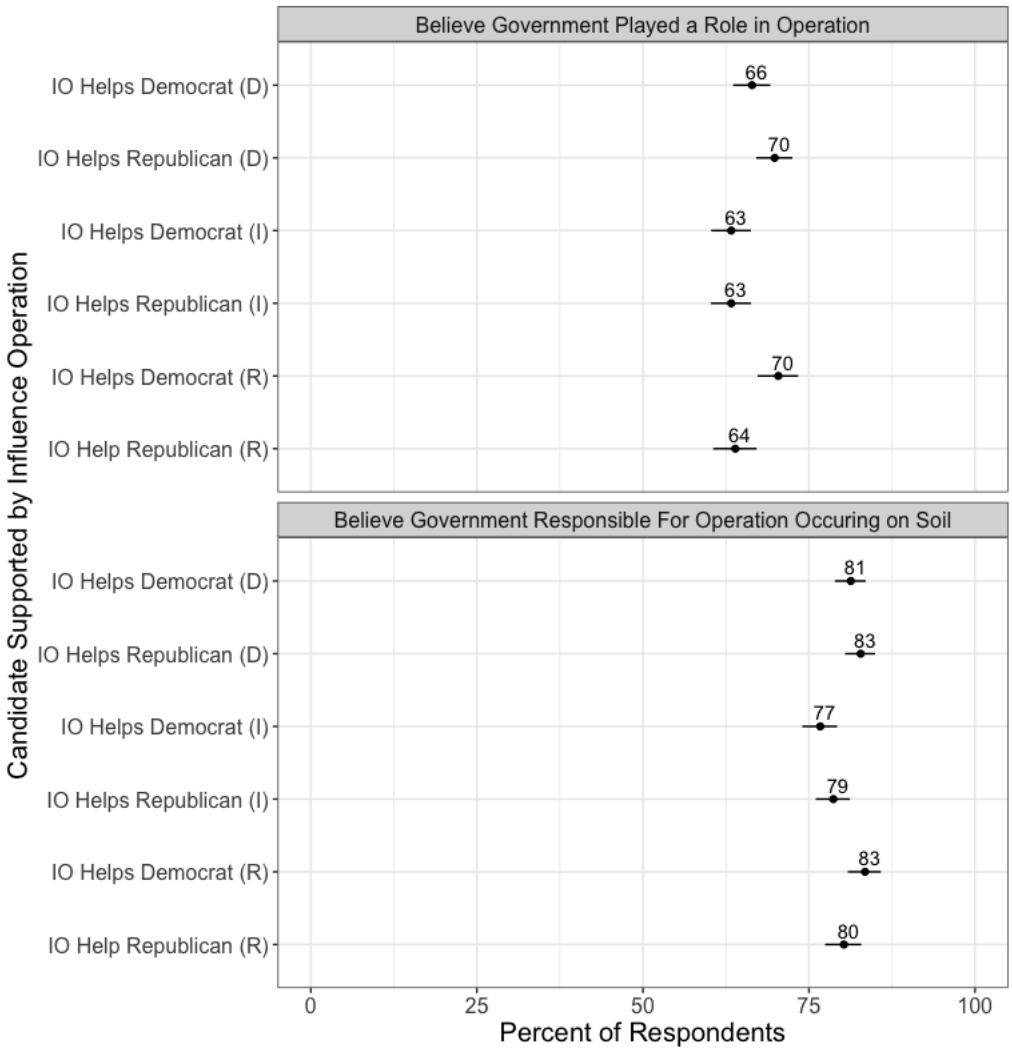
day activities of the U.S. government. In Figure 15 in the Appendix, I follow the same analytical strategy as in Figure 9, but look instead at the percent of respondents who believe the foreign government monitors their online activities or influences the U.S. government's day-to-day activities. I find that implausible deniability (compared to plausible deniability) has a 4-percentage point effect on respondents' belief that the country monitors their online activities "A great deal" or "Quite a bit" and a 5-percentage point effect that the foreign government influences the day-to-day activities of the U.S. government when the perpetrator is an ally. Furthermore, when I recode to look at those who believe the country monitors online activities or influence day-to-day activities of the U.S. government "A great deal," "Quite a bit," or "Somewhat," (including the middle response option), the difference between plausible and implausible deniability treatments for allies jumps to 10 percentage points. I do not, however, find an effect for adversaries. These findings suggest that Russia and China may have already developed reputations for foreign interference, and thus news of an implausibly deniable influence operation does not change views of those countries. However, when respondents learn of operations from Japan and the United Kingdom, countries that lack that reputation, implausible deniability does change views of interference beyond the specific operation.

Partisan Double Standards

To test whether Americans exhibit double standards based on who the operation supports, I first split respondents based on whether they identify as Democrats, Republicans or Independents. I examine whether blame and support for retaliation differ for each group based on whether the operation helps the Democratic candidate or the Republican candidate.

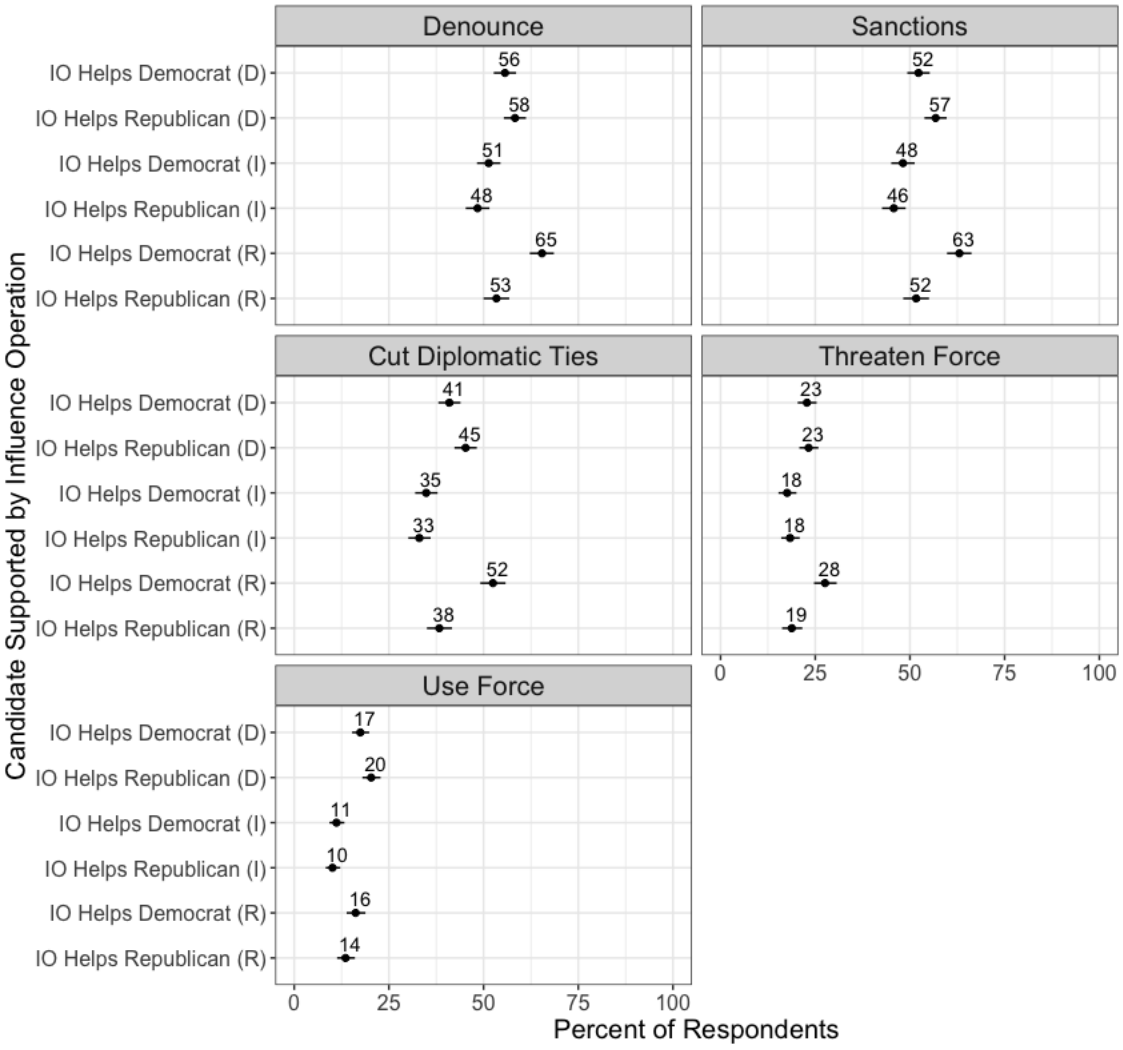
In Figure 11, I plot the percent of respondents who believe the foreign government played a role in the operation and was responsible for the operation occurring on their soil. The axis ticks reflect which candidate the influence operation (IO) helped, and whether the respondents identify as Democrats (D), Republicans (R), or Independents/Other (I). Democrats exhibit a 4-percentage point double standard between when an operation helps the Republican candidate compared to the Democratic candidate on the belief that the government played a role in the operation, and Republicans exhibit a 6-percentage point double standard between when an operation helps the Democratic candidate compared to the Republican one. The double standard is smaller for the belief that a government is responsible for the operation occurring on its soil.

Figure 11: Percent of Respondents Who Blame the Foreign Government



In Figure 12, I plot the percent of respondents who support various forms of retaliation in response to the influence operation. Democrats exhibit a small double standard. The difference in the percent of supporters for denouncing is 2 percentage points when an operation helps the Democrat compared to the Republican, 5 percentage points for sanctions, and 4 percentage points for cutting diplomatic ties. Republicans, however, exhibit a strong double standard of 12 percentage points for denouncing, 11 percentage points for sanctions, 14 percentage points for cutting diplomatic ties and 9 percentage points for threatening the use of force. There is not a strong double standard for using force (2 percentage points), likely due to the floor effect mentioned above.

Figure 12: Percent of Respondents Who Support Retaliation



Conclusion

The results from this study provide several novel insights for understanding plausible deniability and foreign influence operations. In the primary analysis, I found moderate but significant effects for the supposed benefits of plausible deniability; when a foreign government's role in a foreign influence operation was not known, the country was not seen as unfavorably and a smaller percentage of Americans supported retaliatory measures. At the same time, I did not find evidence of the supposed benefits of implausible deniability. Americans were no less paranoid about the foreign government's espionage or influence over U.S. politics and no more fearful of its disinformation campaigns when the foreign government conducted or sponsored the operation compared to when its role was not known. The primary analysis

suggests that states benefit from outsourcing to third-party firms, because third-party firms offer benefits of plausible deniability and getting caught red-handed may not have the benefits some believe. Within the context of my broader dissertation, this study supports the logic that states may pursue the covert route to avoid political repercussions.

In addition to the primary analysis, I conducted a secondary analysis to examine how identity of the country of origin and partisanship affect Americans' assignments of blame and support for retaliation. Americans were far more likely to believe that a country played a role when it is an adversary compared to an ally and they were more likely to support retaliation against an adversary than an ally. In fact, plausible deniability did not have an effect on belief that an enemy was involved in an operation or on support for retaliation, while it did have an effect on both outcomes for allies. These findings may help to explain why Russia continues to wage influence operations even after it has repeatedly been caught; at this stage, operations with even loose links will be attributed to Russia so the benefits of plausible deniability are weak.

I also investigated the role of partisanship in assignment of blame and support for retaliation. Respondents did not exhibit a strong double standard on assignment of blame, but Republicans exhibited a notable double standard on support for retaliation. When an operation supported a Democrat, Republicans were significantly more supportive of retaliation compared to when an operation supported a Republican. Although the partisan double standard is weaker than in Tomz and Weeks 2020, my results provide further evidence that foreign influence operations that support one candidate and disparage another can amplify existing political tensions through disparate support for retaliation.

The findings from this study provide the basis for a number of potential research projects to come. First, this study provides evidence that some governments benefit from plausible deniability while others do not. Further research could extend this analysis to examine a broader range of countries. The finding that Russia and China do not benefit from plausible deniability may be unique to these two countries, given that Americans may perceive that they wage frequent influence operations while not holding these views for other low-favorability countries. Second, future experiments can address other types of uncertainty. Influence operations that are unmasked are increasingly waged by domestic political actors—even actors vying for influence within their own parties. Additional experiments could demonstrate how citizens in different countries react to domestic versus foreign operations. This would test the effects of plausible

deniability when partisanship is less likely to drive results. Third, I found that many respondents believe that fake accounts likely had an effect on election results. Further research should explore how governments in democratic states might normalize citizens to the fact that such fake accounts exist, so that news of them near an election does not produce disproportionate views of their impact. Finally, this paper relied on survey experiments to test the supposed benefits of plausible and implausible deniability. In the real world, Facebook and Twitter have made dozens of attributions of influence operations in the last few years alone. Observational studies could investigate the impact of these attributions: What is the effect on marketing firms that run a disinformation campaign when a platform publicly attributes a campaign to them? Such a project might begin by constructing a list of private firms to which influence operations have been publicly attributed, and then study the factors that determine whether those attributions affect the perpetrator.

Appendix

Sample Information

To gather the sample, I deployed a survey with Lucid Theorem. Lucid Theorem is a panel provider that creates samples designed to approximate the U.S. population. They quota sample based on age, gender, ethnicity and region. Many recent papers in Political Science and misinformation studies use Lucid to construct samples.[97]

After respondents provide informed consent, I presented them with a matrix of questions: three questions were standard ideological questions (Hawk, Internationalism, Trust), while the other two questions were attention checks. The attention checks were simple. I asked respondents to select the "Disagree" option and whether the year 1910 came before the year 1920. Those who failed the attention checks were speed readers, and were excluded from the remainder of the survey.

After collecting the data, I analyzed the number of people who failed at least one of the two attention checks: 2,174 people failed one of the two attention checks and were excluded from participating in the survey, while 7,792 passed the attention checks. The 7,792 people who passed the attention checks were used throughout the paper.

To ensure that the sample was evenly distributed across treatment groups, I checked the number of respondents in the sample assigned to each group. Between 1,507 and 1,509 respondents were assigned to each of the five attribution conditions, and 1,885 or 1,886 were assigned to each of the four Country treatments.

The sample was 46.9% male, 52.4% female, and .7% who did not identify as male or female. In terms of race, 77.2% of the sample identify as White, 12.4% as Black or African American, 11.5% Hispanic, and 4.8% Asian. The mean respondent age was 46.3 years, and 36.1% of the sample completed a 4-year or postgraduate degree. In terms of party affiliation, 28.6% of respondents think of themselves as Republicans and 36.5% as Democrats.

[97] For a few examples, Bode and Vraga 2021; Robinson 2021; Stecula and Pickup 2021.

Testing Heterogenous Treatment Effects for Partisanship

Hypothesis 11 posits heterogenous treatment effects, wherein plausible deniability has a stronger effect on the belief that a foreign government was involved in an operation or is responsible for it occurring on their soil when the operation helps the opposition candidate compared to one's own candidate. However, due to space constraints, I did not present results for the hypothesis in the main paper.

To conduct this analysis, I drop Independents from the dataset. I create groups for Favored Candidate (Democratic respondents when an operation helps a Democrat, and Republican respondents when an operation helps a Republican) and Opposition (Democratic respondents when an operation helps a Republican, and Republican respondents when an operation helps Democrats). I then look at whether the effect of plausible versus implausible deniability is bigger for Favored Candidate or Opposition.

As I show in the Figure 13, I do not find significant double standards on assignment of blame. However, as shown in Figure 14, there are significant double standards when it comes to support for retaliating. When an operation helps a respondent's favored candidate, a similar percent of respondents support denouncing the country, sanctioning the country, and cutting diplomatic ties with the country when the operation is plausibly deniable or when a government's involvement is known. By contrast, when an operation helps the opposition, the difference in percent of respondents who support retaliating when an operation is implausibly deniable compared to when it is plausibly deniable is 9% for denouncing (66% compared to 57%), 9% for sanctions (64% compared to 55%), and 7% for cutting diplomatic ties (52% compared to 45%). The double standard is small for more severe forms of using force, where a floor effect is likely at play.

Figure 13: Percent of Respondents Who Blame Foreign Government

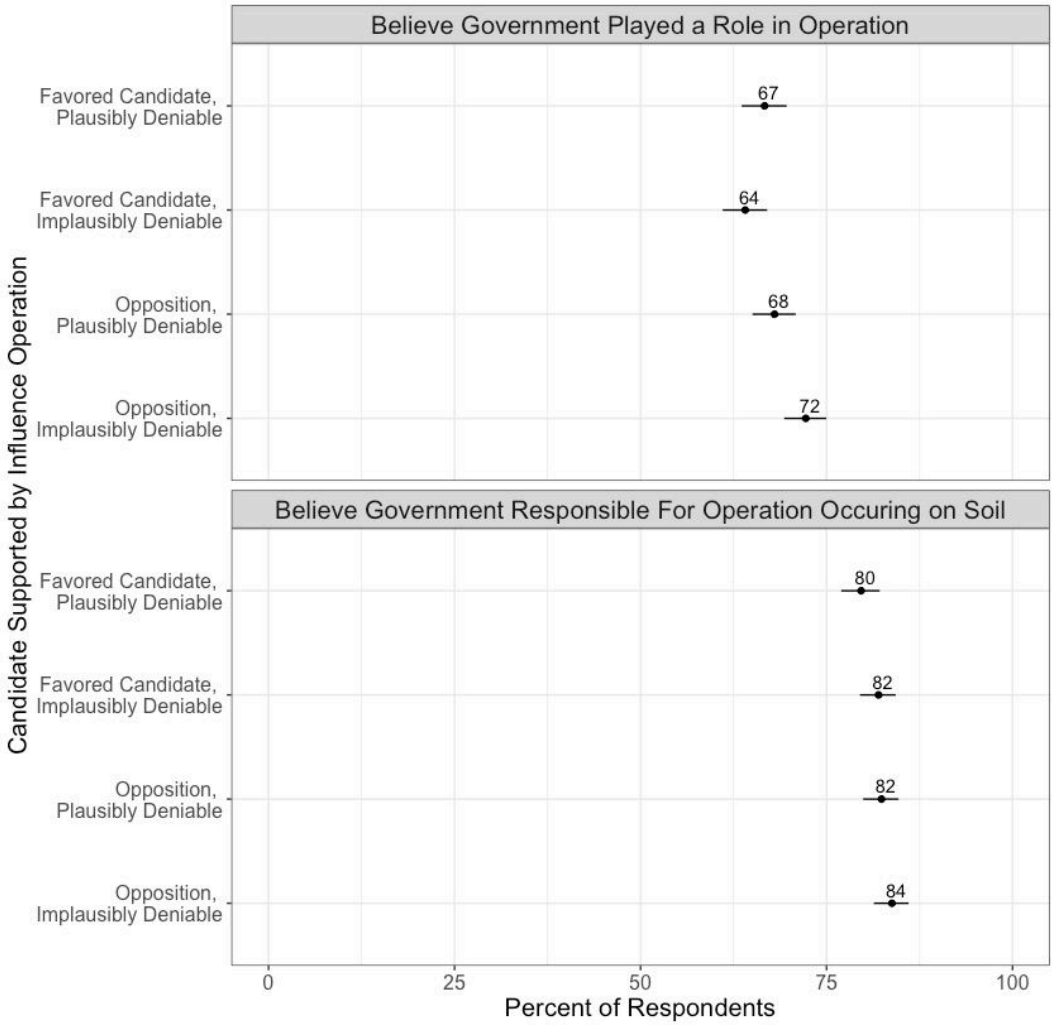
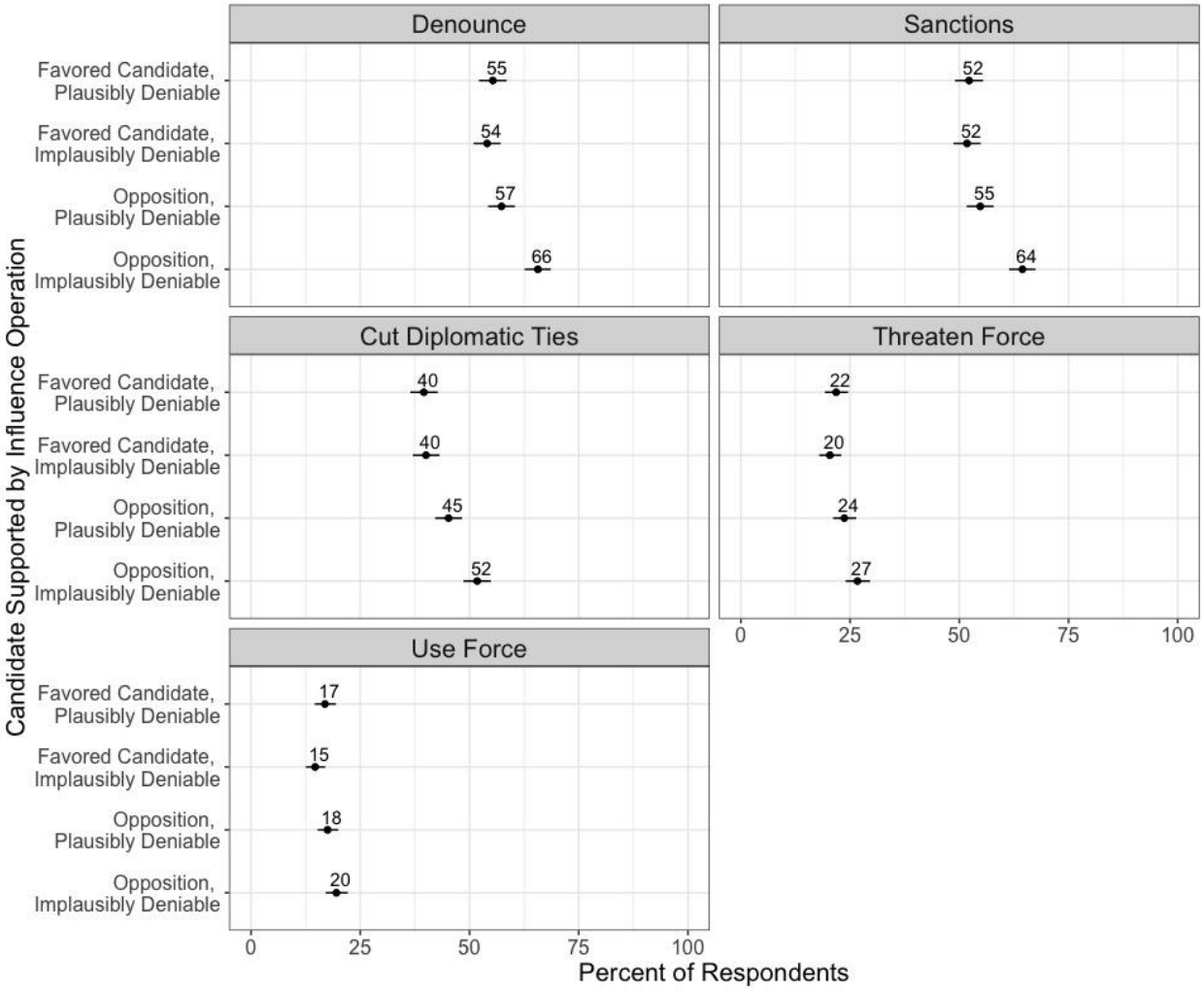


Figure 14: Percent of Respondents Who Support Retaliation



Testing Heterogenous Treatment Effects of Plausible Deniability and Country of Origin

In the main text, I found null effects in Figure 6. When operations were implausibly deniable, respondents were no more likely to believe that the country of origin monitors their online activities or influences the day-to-day activities of the U.S. government compared to when operations were plausibly deniable. This null finding hides variation in the effect of plausible deniability on these dependent variables based on the identity of the country of origin of an operation.

Figure 15: Percent of Respondents Who Believe That…

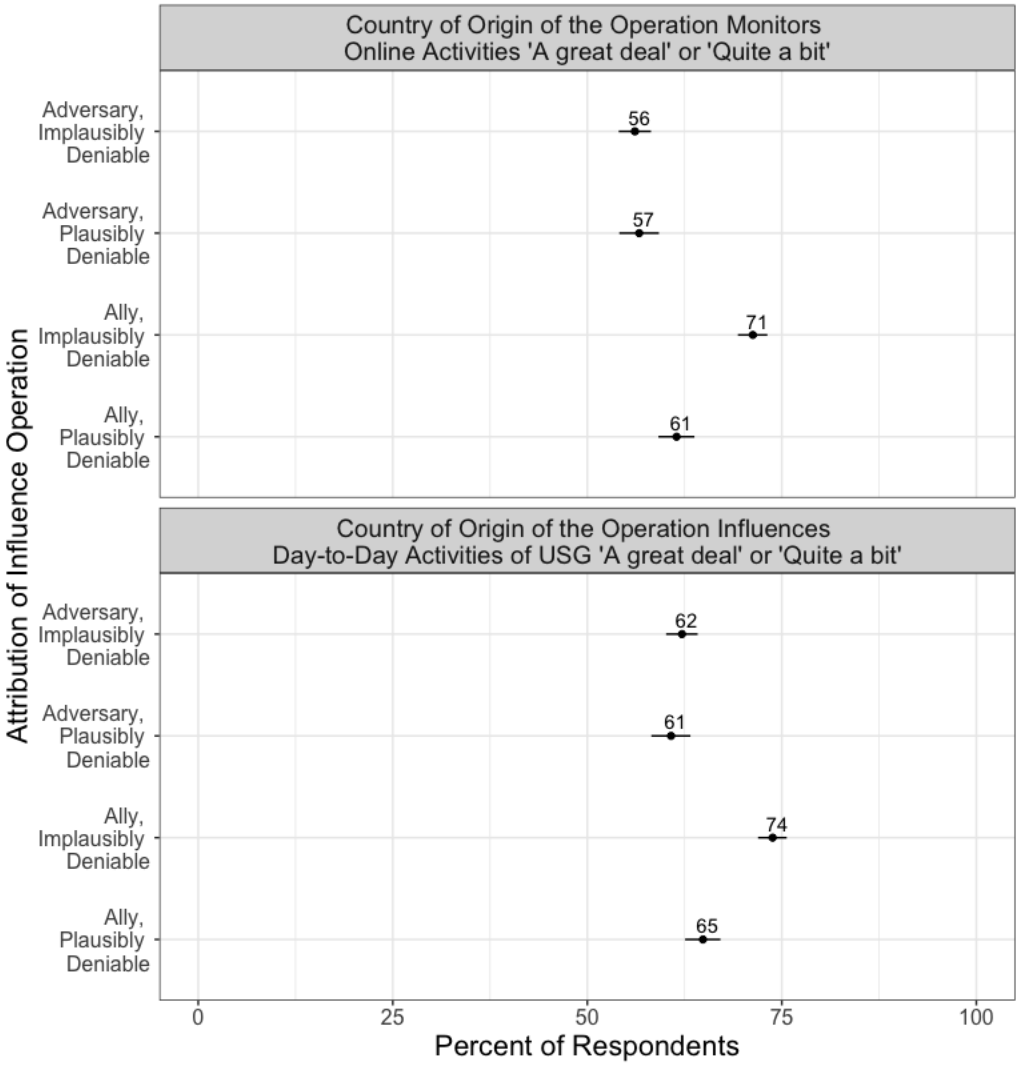
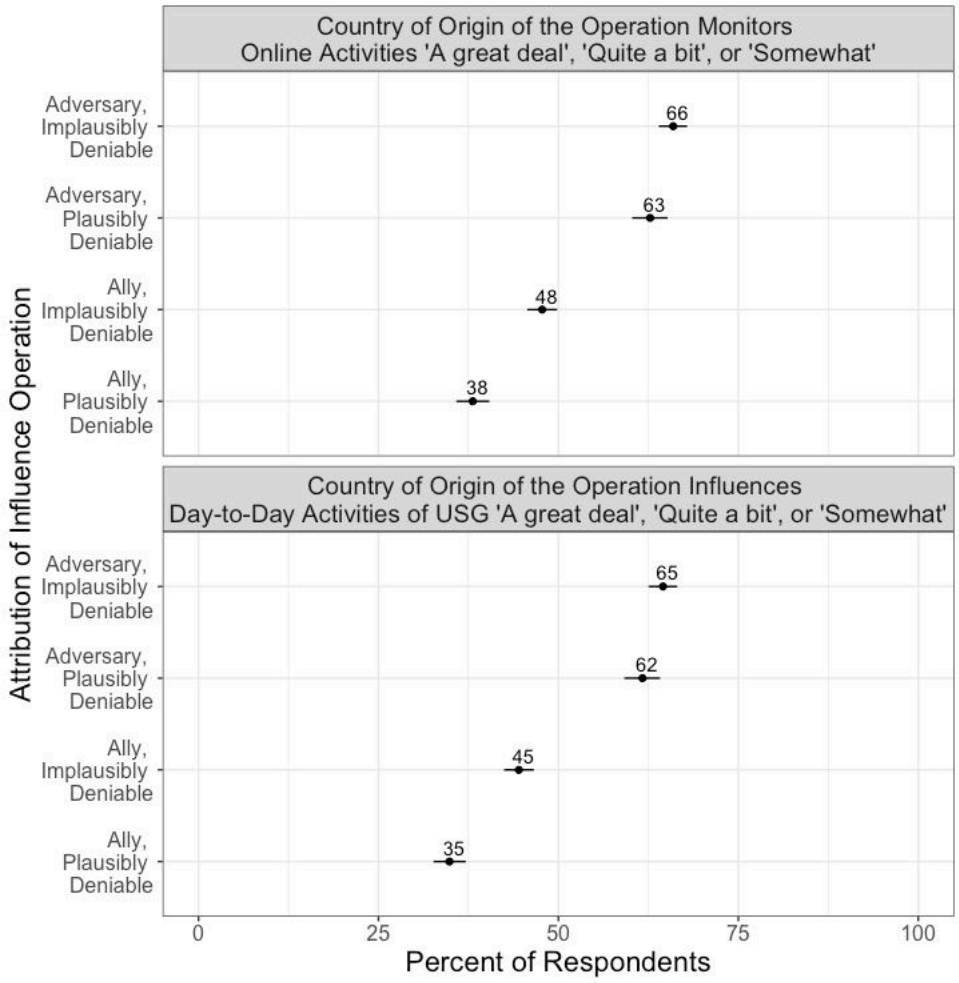


Figure 16: Percent of Respondents Who Believe That…



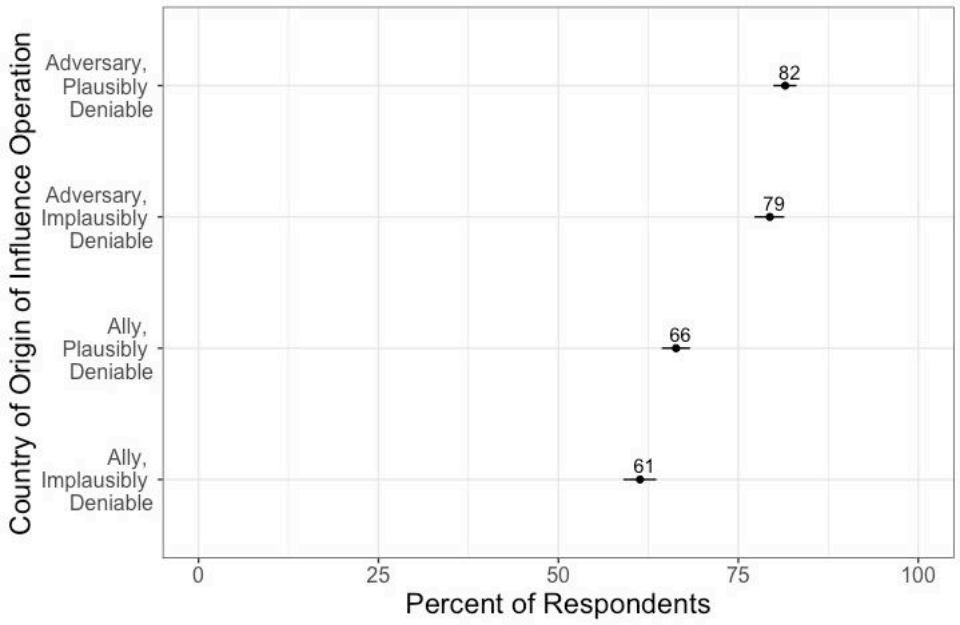
Believability of Scenario

To examine the believability of the vignette, I asked respondents the question: "How believable did you find this scenario?" This was the last question of the survey, and therefore did not contaminate any of the questions presented in the paper or in this Appendix. Respondents were offered four answer options: "Very believable," "Somewhat believable," "Somewhat unbelievable," and "Not believable at all."

In Figure 17, I plot the percent of respondents who reported that the scenario was "Very believable" or "Somewhat believable" by whether the country was an ally or adversary and whether they received a plausibly deniable or implausibly deniable treatment. There is a substantial effect of a country's identity on believability: 82% and 79% of respondents believed

the scenario was believable when it was a plausibly deniable operation by an adversary, and an implausibly deniable operation by an adversary, respectively. However, 66% of respondents believed the scenario was believable when it was a plausibly deniable operation by an ally, and 61% when it was an implausibly deniable operation by an ally.

Figure 17: Percent of Respondents Who Report the Scenario was "Very believable" or "Somewhat believable"



Reproductions of Main Text Results with Pooled or Individual Treatments

In this section, as I registered in the pre-analysis plan, I reproduce graphs from the main text with pooled or non-pooled treatments. In cases where I presented treatment groups independently in the main text, in the Appendix I present results for pooled results. For example, if I presented each plausible/implausible deniable treatment in a main text graph, here I reproduce results for two groups: treatments that were plausibly deniable and treatments that were implausibly deniable. I follow the same strategy for either grouping or ungrouping countries based on whether they are a favorable ally or unfavorable adversary. The conclusions presented in the main text are consistent with results presented from the alternate empirical strategy in the Appendix.

Figure 18: Percent of Respondents Who Support Retaliation, by Binary Plausibility

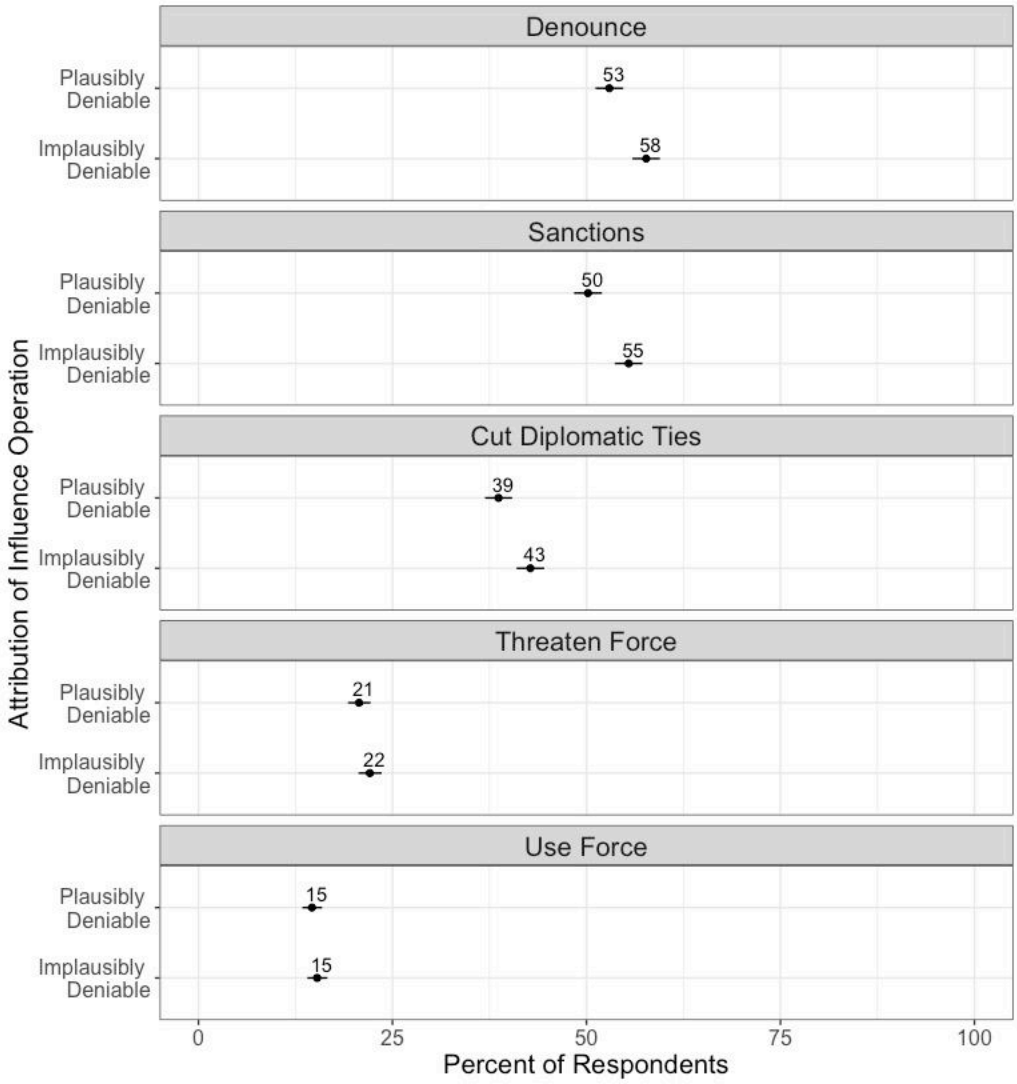


Figure 19: Percent of Respondents Who Would View Perpetrator Negatively, by Binary Plausibility

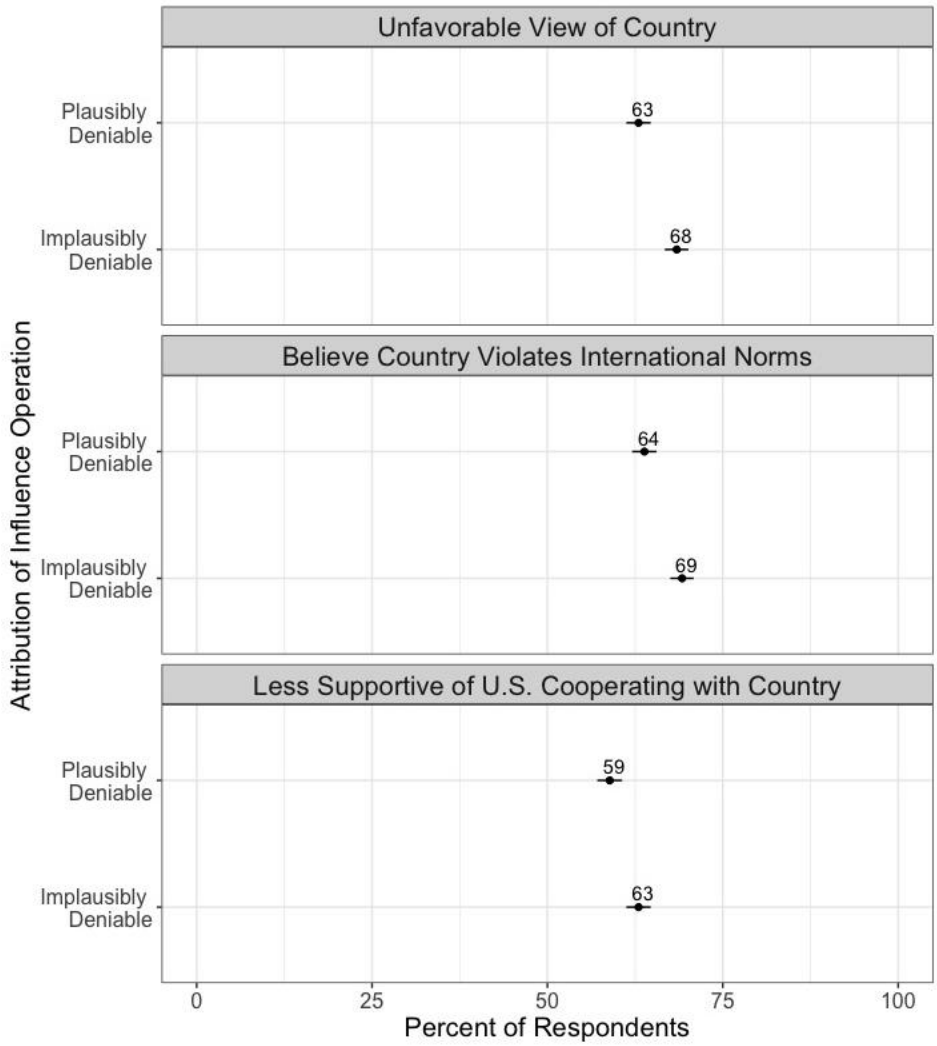


Figure 20: Percent of Respondents Who Believe Operation Had An Effect, by Binary Plausibility

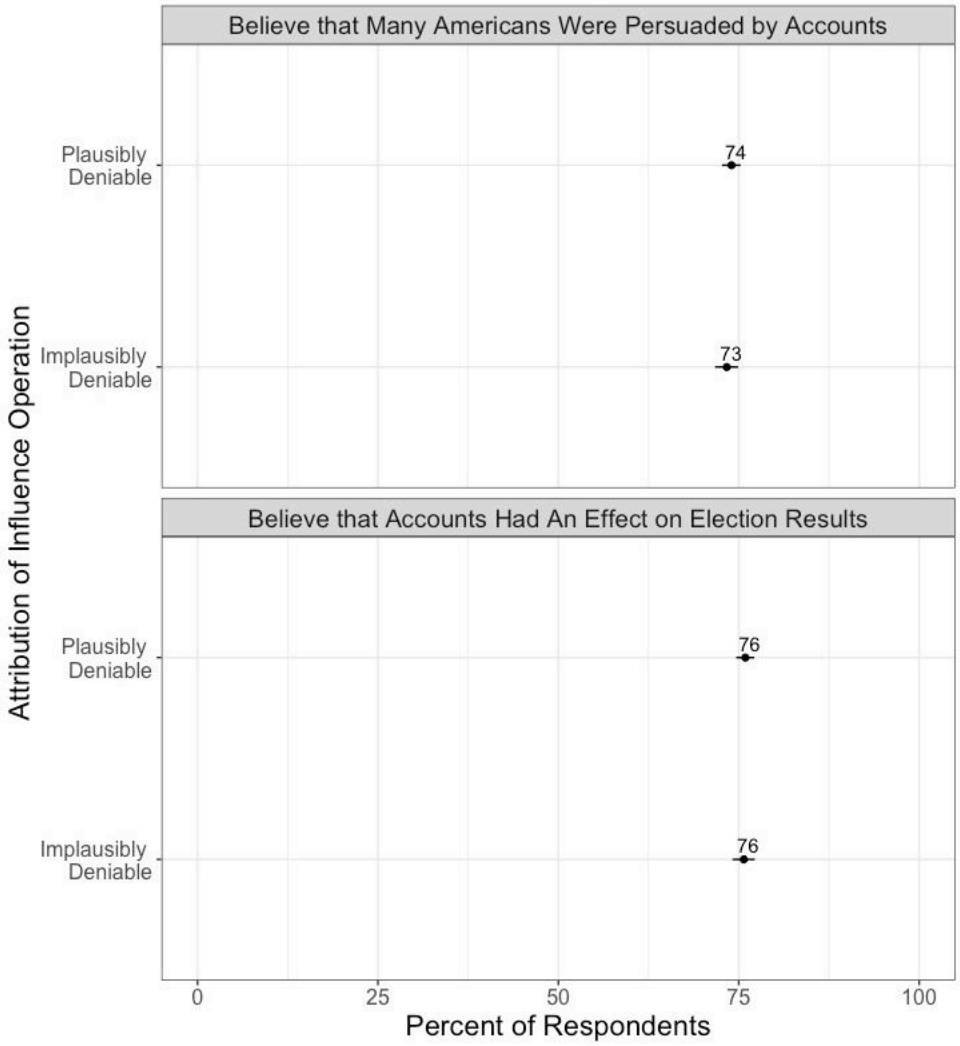


Figure 21: Percent of Respondents Who Believe Perpetrator Monitors Activity or Influences USG, by Binary Plausibility

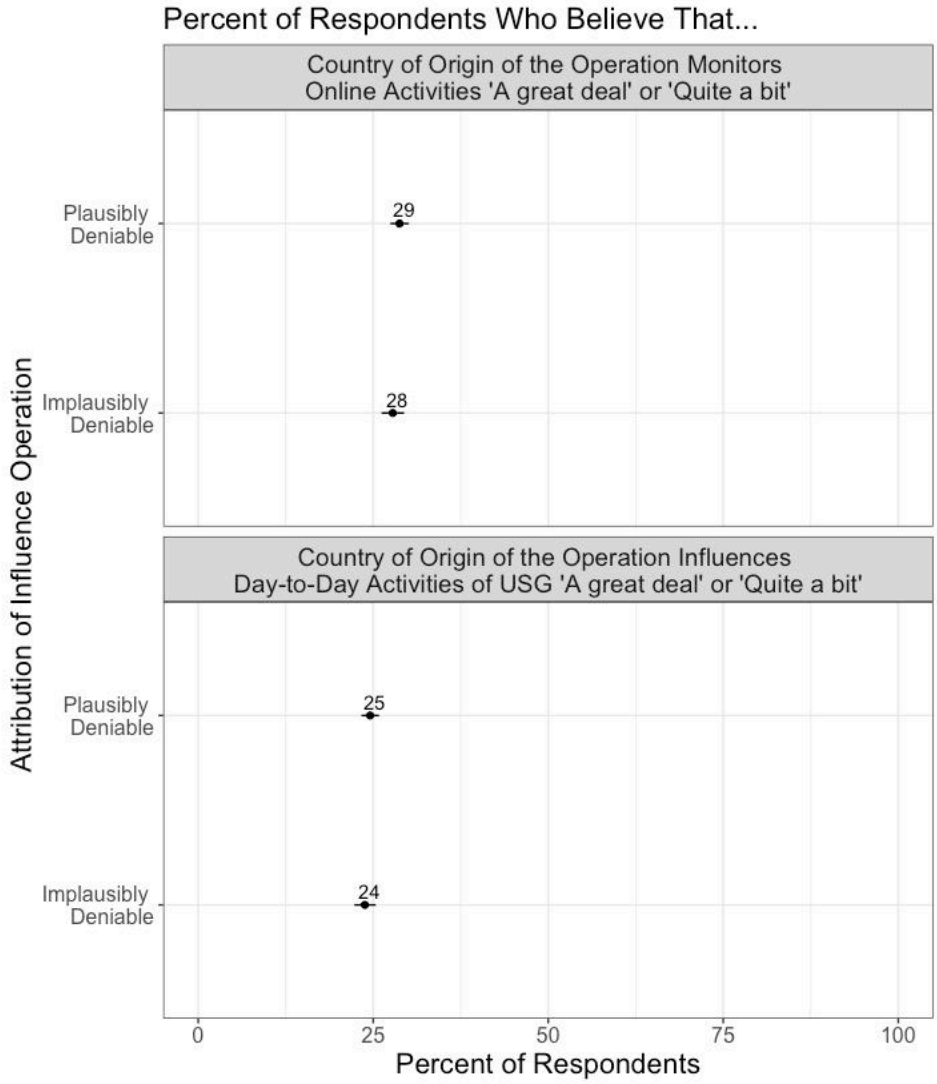


Figure 22: Percent of Respondents Who Believe Government Involved, by Binary Plausibility

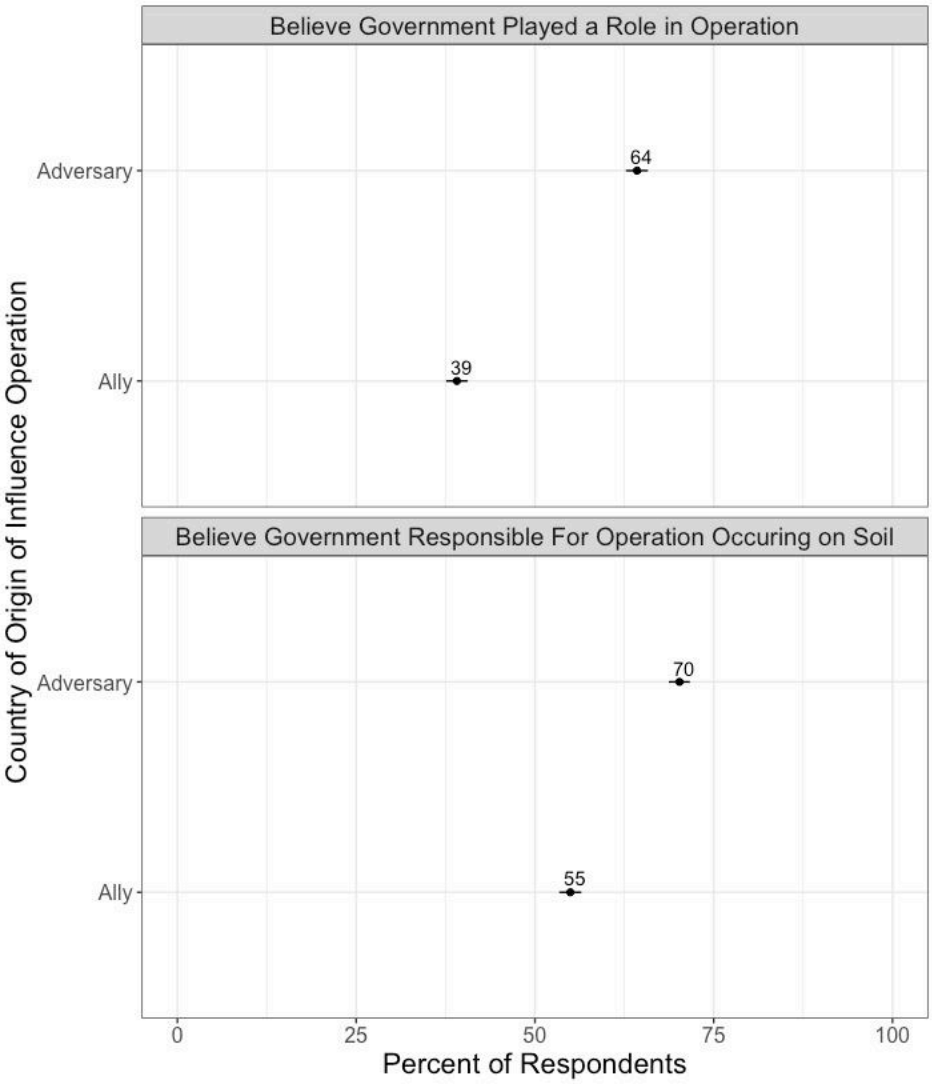
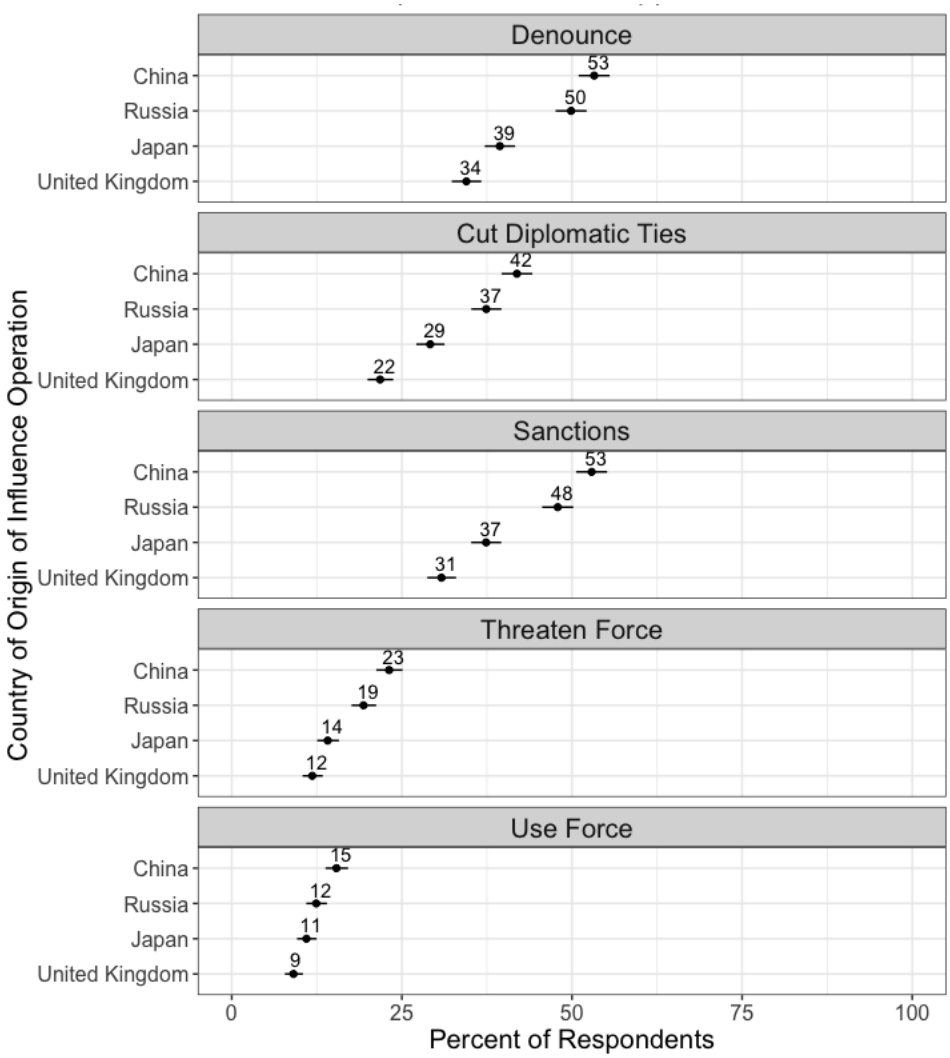


Figure 23: Percent of Respondents Who Support Retaliation, Full Country



Reproductions of Main Text Primary Results With Attentive Readers

After the vignette, I included attention checks, including:

- What year did the operation occur?
- Who was the operation designed to benefit?

Of the 7,792 respondents, 5,908 correctly answered both attention checks. In the graphs below, I reproduce all graphs from the primarily analysis section of the paper including only those 75.8% of respondents who attentively read the survey. As in the paper, I find evidence of the benefits of plausible deniability and a null finding for evidence of implausible deniability.

Figure 24: Percent of Respondents Who Support Retaliation, Attentive Readers

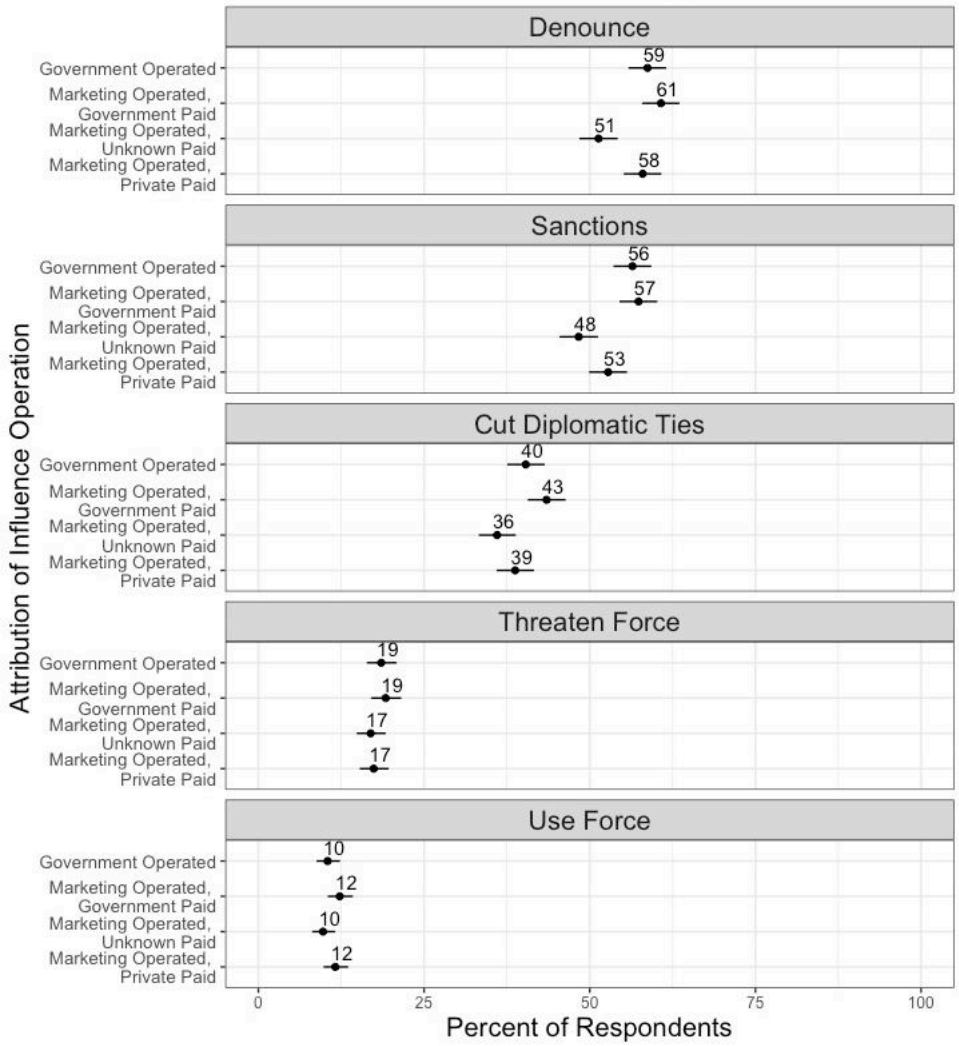


Figure 25: Percent of Respondents Who Hold Unfavorable View of Country, Attentive Readers



Figure 26: Percent of Respondents Who Believe Operation Had An Effect, Attentive Readers

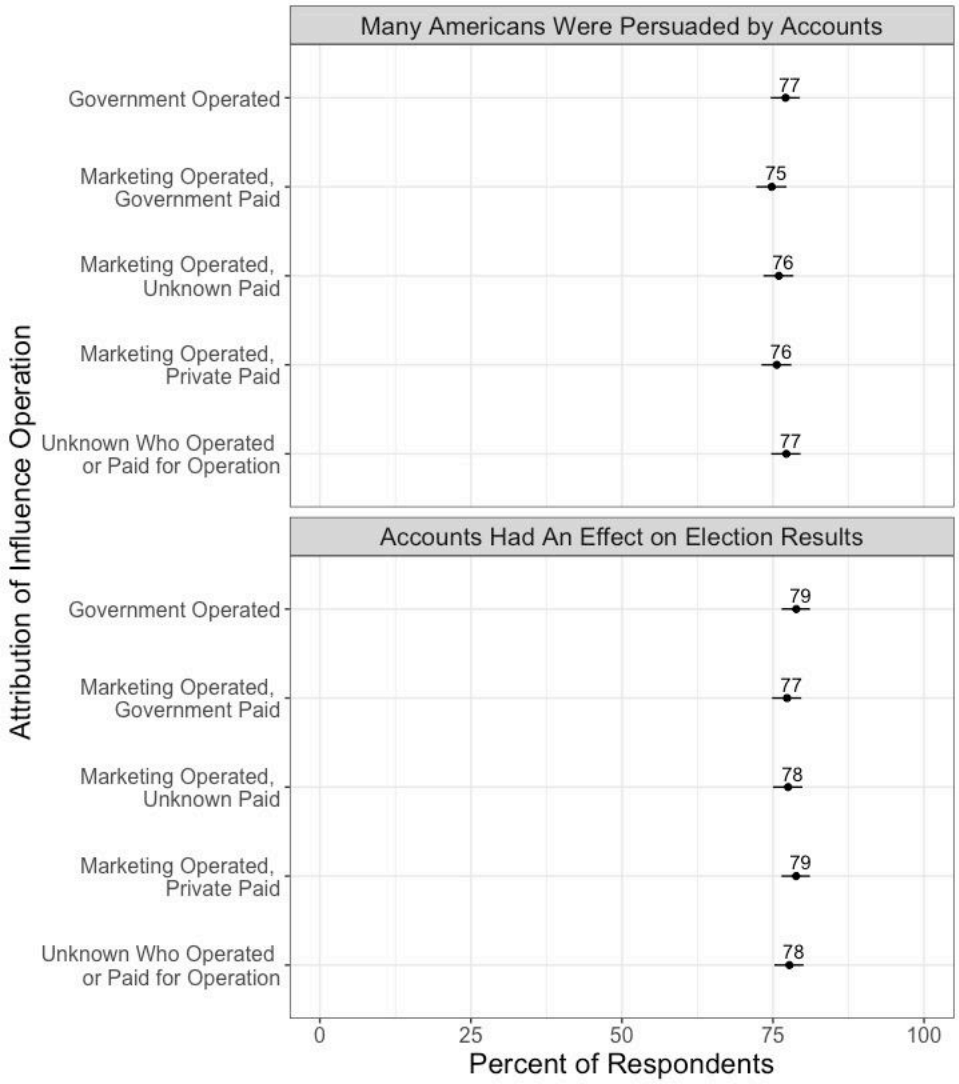
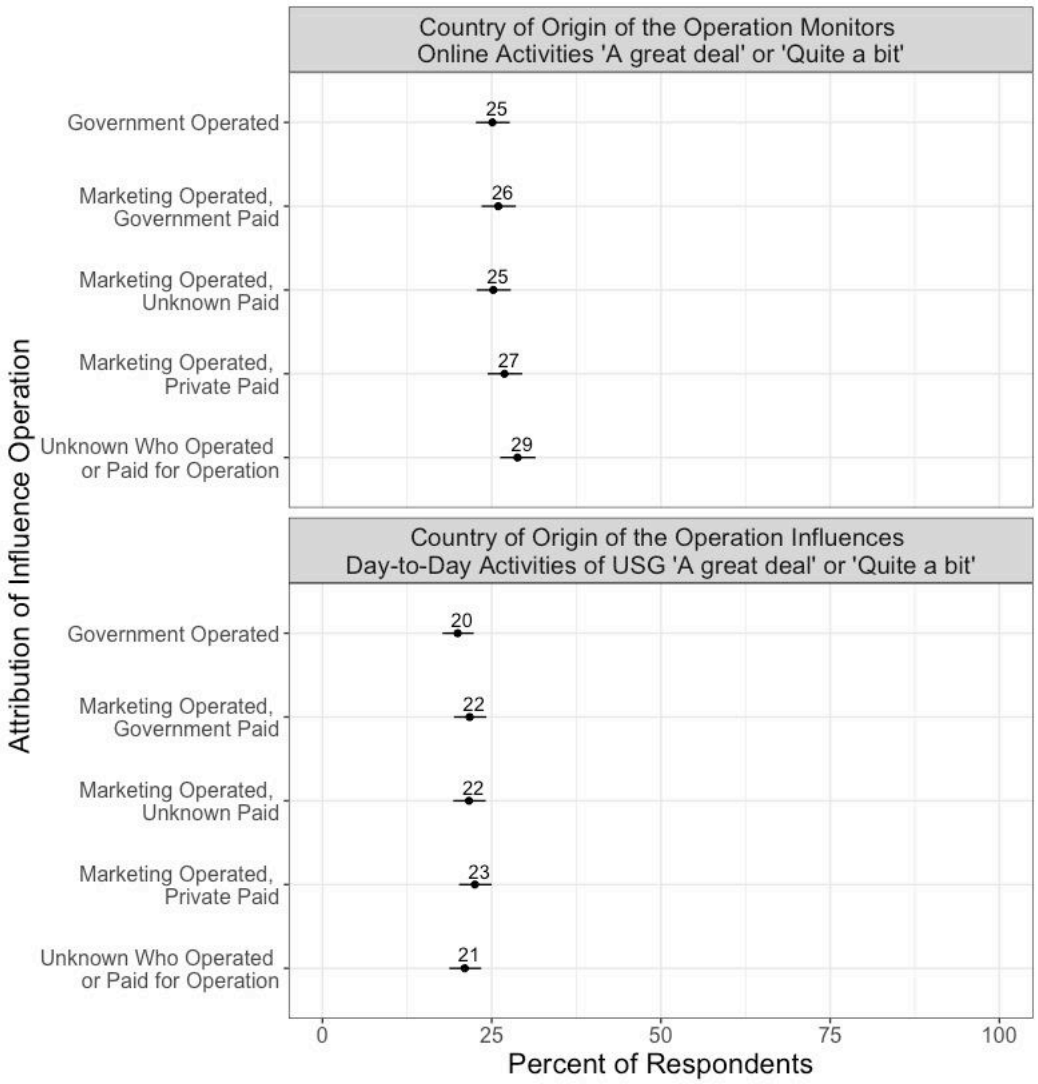


Figure 27: Percent of Respondents Who Believe Country Monitors Activity or Influences USG, Attentive Readers



Survey Questionnaire

Foreign Interference

Start of Block: Introduction

Q1 We appreciate your interest in participating in this online survey. Please read through this information before clicking the "next" button at the bottom of the page if you agree to participate.

You will be asked some questions about yourself and then asked to provide responses to a hypothetical scenario about American politics. This should take about 10 minutes. No background knowledge is required and we will not collect any data that could directly identify you. The aggregated results may be published in an article or book.

Please note that your participation is voluntary, and that you may withdraw at any point by closing the browser. However, we are only able to pay participants who complete the survey. You may only participate in this survey if you are 18 years of age or over.

This project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee (SSH_DPIR_C1A_21_014). If you have a concern about any aspect of this study, please email Janina Dill [janina.dill@politics.ox.ac.uk]. If you remain unhappy or wish to make a formal complaint, please contact ethics@socsci.ox.ac.uk.

If you have read the information above and agree to participate with the understanding that the data you submit will be processed accordingly, click the arrow below to get started.

End of Block: Introduction

Start of Block: Default Question Block

Q2 We are first going to ask you some background questions.

End of Block: Default Question Block

Start of Block: Attention Check

Q3 To what extent do you agree or disagree with the following statements?

	Disagree strongly (1)	Disagree (2)	Neither agree nor disagree (3)	Agree (4)	Agree strongly (5)
The year 1910 came before the year 1920 (1)	○	○	○	○	○
The use of military force only makes problems worse (2)	○	○	○	○	○
The United States needs to play an active role in solving conflicts around the world (3)	○	○	○	○	○
Please click the "Disagree" response (4)	○	○	○	○	○
The best way to ensure world peace is through American military strength (5)	○	○	○	○	○

End of Block: Attention Check

Start of Block: Demographics

Q4 Do you describe yourself as a man, a woman, or in some other way?

- Man (1)
- Woman (2)
- Other (3) _____

Q5 Which of the following best describes your race? (mark all that apply)

- White (1)
- Black or African American (2)
- American Indian or Alaska Native (3)
- Asian (4)
- Native Hawaiian or Pacific Islander (5)
- Other (please specify) (6) _____

Q6 Are you of Hispanic or Latino origin?

- No (1)
- Yes, Mexican, Mexican American, Chicano (2)
- Yes, Puerto Rican (3)
- Yes, Cuban (4)
- Yes, another Hispanic, Latino, or Spanish origin (please specify) (5)
  _____

Q7 In what state do you currently reside?

▼ Alabama (1) ... I do not reside in the United States (53)

Q8 How old are you?

▼ 18 (1) ... 105 (88)

Q9 What is the highest level of school you have completed?

- Did not graduate from high school (1)
- High school graduate (2)
- Some college, but no degree (yet) (3)
- 2-year college degree (4)
- 4-year college degree (5)
- Postgraduate degree (MA, MBA, MD, JD, PhD, etc) (6)

Page Break

Q10 Generally speaking, do you think of yourself as a ...

- Republican (1)
- Democrat (2)
- Independent (3)
- Another party, please specify (4) _____
- No preference (5)

Display This Question:
If Generally speaking, do you think of yourself as a ... != Republican
And Generally speaking, do you think of yourself as a ... != Democrat

Q11 Do you think of yourself as closer to the ...

- Republican Party (1)
- Democratic Party (2)
- Neither party (3)

Display This Question: If Generally speaking, do you think of yourself as a ... = Democrat

Q12 Would you call yourself a ...
- Strong Democrat (1)
- Not very strong Democrat (2)

Display This Question: If Generally speaking, do you think of yourself as a ... = Republican

Q13 Would you call yourself a ...
- Strong Republican (1)
- Not very strong Republican (2)

Q14 In general, do you think of yourself as ...
- Extremely liberal (1)
- Liberal (2)
- Slightly liberal (3)
- Moderate, middle of the road (4)
- Slightly conservative (5)
- Conservative (6)
- Extremely conservative (7)

Page Break

Q15 How often do you read the news?
- Daily (1)
- A few times a week (2)
- Once a week (3)
- Once a month (4)
- Never (5)

Q16 How frequently do you use social media?
- Several times a day (1)
- About once a day (2)
- A few days a week (3)
- Every few weeks (4)
- Less often (5)

Q17 Overall, what effect would you say social media has on the way things are going in this country?
- Mostly negative (1)
- Neither positive nor negative (2)
- Mostly positive (3)

Page Break

Q18 Generally speaking, would you say that the U.S. can trust other nations, or that the U.S. should be very careful in dealing with other nations?

- The U.S. can trust other nations (1)
- The U.S. should be very careful in dealing with other nations (2)

Q19 Please tell me if you have a very unfavorable, somewhat unfavorable, somewhat favorable, or very favorable view of …

	Very unfavorable (1)	Somewhat unfavorable (2)	Somewhat favorable (3)	Very favorable (4)
China (China_PreT)	○	○	○	○
Russia (Russia_PreT)	○	○	○	○
The United Kingdom (UK_PreT)	○	○	○	○
Japan (Japan_PreT)	○	○	○	○

Page Break

Q20 Do you approve or disapprove of how Joe Biden is handling his job as president?

- Approve (1)
- Disapprove (2)
- Don't know (3)

Q21 Do you approve or disapprove of how Donald Trump handled his job as president?

- Approve (1)
- Disapprove (2)
- Don't know (3)

Q22 If you were to vote in the 2024 U.S. election, would you most likely vote for the ...

- Democratic candidate (1)
- Republican candidate (2)
- Neither the Democratic nor the Republican candidate (3)
- I do not vote (4)

End of Block: Demographics

Start of Block: Block 6

Q23 Next, we will describe a hypothetical situation that could take place in the future. Please read the description carefully. After you have read about the situation, we will ask for your opinions.

End of Block: Block 6

Start of Block: Vignette

Q24 Please read the following scenario carefully:

In 2024, during the U.S. Presidential election campaign, a network of thousands of fake accounts on social media pretended to be real Americans and shared content that favored the ${e://Field/Candidate} and criticized the ${e://Field/Opposition}.

The U.S. intelligence community ${e://Field/Attribution}.

${e://Field/Denial}

Q25 Before continuing, we need to make sure you read this information carefully.

Q26 In which year does the campaign take place?

- 2016 (1)
- 2020 (2)
- 2024 (3)
- 2028 (4)

Q27 What was the social media campaign designed to do?

- Help the ${e://Field/Opposition} and hurt the ${e://Field/Candidate} (1)
- Help the ${e://Field/Candidate} and hurt the ${e://Field/Opposition} (2)
- Hurt both the ${e://Field/Candidate} and the ${e://Field/Opposition} (3)
- No information given (4)

Display This Question: If Attribution != does not know who operated the accounts or paid for the operation

Q28 According to the U.S. intelligence community, which of the following actors operated these fake social media accounts?

- The ${e://Field/Country2} government (1)
- A marketing firm in ${e://Field/Country} (2)
- Teenagers in ${e://Field/Country} (3)
- They do not know who operated the accounts (4)

End of Block: Vignette

Start of Block: Block 5

Display This Question: If Attribution != does not know who operated the accounts or paid for the operation

Q29 Do you believe that the ${e://Field/Country2} government played a role in the operation?

- Definitely (1)
- Probably (2)
- Probably not (3)
- Definitely not (4)

Display This Question: If Attribution != does not know who operated the accounts or paid for the operation

Q30 To what extent is the ${e://Field/Country2} government responsible for the operation occurring on ${e://Field/Country2} soil?

- Entirely responsible (1)
- Somewhat responsible (2)
- Probably not responsible (3)
- Not at all responsible (4)

Page Break

Q31 Here is the information again for your reference:

In 2024, during the U.S. Presidential election campaign, a network of thousands of fake accounts on social media pretended to be real Americans and shared content that favored the ${e://Field/Candidate} and criticized the ${e://Field/Opposition}.

The U.S. intelligence community ${e://Field/Attribution}.

${e://Field/Denial}

Q32 If the 2024 election happened just as we described, would you agree or disagree with each of the following statements?

	Strongly disagree (1)	Somewhat disagree (2)	Somewhat agree (3)	Strongly agree (4)
Many Americans were persuaded by information coming from these fake social media accounts (1)	○	○	○	○
These fake social media accounts likely had an effect on election results (2)	○	○	○	○

Q33 If you had to guess, would you say that these fake social media accounts ...

- Gave the ${e://Field/Candidate} a big advantage (1)
- Gave the ${e://Field/Candidate} a small advantage (2)
- Didn't give either side an advantage (3)
- Gave the ${e://Field/Opposition} a small advantage (4)
- Gave the ${e://Field/Opposition} a big advantage (5)

Q34 Please type a few sentences explaining your answer.

_____
_____
_____
_____
_____

Page Break

Q46 Here is the information again for your reference:

In 2024, during the U.S. Presidential election campaign, a network of thousands of fake accounts on social media pretended to be real Americans and shared content that favored the ${e://Field/Candidate} and criticized the ${e://Field/Opposition}.

The U.S. intelligence community ${e://Field/Attribution}.

${e://Field/Denial}

Display This Question: If Attribution != does not know who operated the accounts or paid for the operation

Q35 If the 2024 election happened just as we described, which policies by the U.S. government would you support or oppose?

	Strongly oppose (1)	Somewhat oppose (2)	Somewhat support (3)	Strongly support (4)
Do nothing	○	○	○	○
Publicly denounce ${e://Field/Country}	○	○	○	○
Cut off diplomatic relations with ${e://Field/Country}	○	○	○	○
Impose economic sanctions on ${e://Field/Country}	○	○	○	○
Threaten to use military force against ${e://Field/Country}	○	○	○	○
Launch a military strike against ${e://Field/Country}	○	○	○	○

Display This Question: If Attribution != does not know who operated the accounts or paid for the operation

Q36 Which of the following policies would you support the most? (Please choose only one.)

- Do nothing (1)
- Publicly denounce ${e://Field/Country} (7)
- Cut off diplomatic relations with ${e://Field/Country} (2)
- Impose economic sanctions on ${e://Field/Country} (3)
- Threaten to use military force against ${e://Field/Country} (4)
- Launch a military strike against ${e://Field/Country} (5)

Display This Question: If Attribution != does not know who operated the accounts or paid for the operation

Q37 Please type a few sentences explaining your answer.

_____
_____
_____
_____
_____

Page Break

Display This Question: If Attribution != does not know who operated the accounts or paid for the operation

Q38 If the 2024 election happened just as we described, would you agree or disagree with each of the following statements?

	Strongly disagree (1)	Somewhat disagree (2)	Somewhat agree (3)	Strongly agree (4)
I would have an unfavorable view of ${e://Field/Country} (1)	○	○	○	○
I would view ${e://Field/Country} as a country that violates international norms (2)	○	○	○	○
I would be less supportive of the United States cooperating with ${e://Field/Country} in the future (3)	○	○	○	○

Page Break

Display This Question: If Attribution != does not know who operated the accounts or paid for the operation

Q39 If the 2024 election happened just as we described, would you agree or disagree with each of the following statements?

	Strongly disagree (1)	Somewhat disagree (2)	Somewhat agree (3)	Strongly agree (4)
${e://Field/CountryCap} is a strong country in international politics (1)	○	○	○	○
${e://Field/CountryCap} is a major threat to the United States (2)	○	○	○	○
${e://Field/CountryCap}'s disinformation campaigns are a major problem for the United States (3)	○	○	○	○

Page Break

Q40 If the 2024 election happened just as we described, would you agree or disagree with each of the following statements?

	Strongly disagree (1)	Somewhat disagree (2)	Somewhat agree (3)	Strongly agree (4)
I would trust the results of the election (1)	○	○	○	○
I would be unlikely to vote in future elections (2)	○	○	○	○
I would lose faith in American democracy (3)	○	○	○	○

Page Break

Q43 Please tell me if you have a very unfavorable, somewhat unfavorable, somewhat favorable, or very favorable view of …

	Very unfavorable (1)	Somewhat unfavorable (2)	Somewhat favorable (3)	Very favorable (4)
China (1)	○	○	○	○
Russia (2)	○	○	○	○
The United Kingdom (3)	○	○	○	○
Japan (4)	○	○	○	○

Page Break

Q41 To what extent do you think the ${e://Field/Country2} government is monitoring your online activities?

- A great deal (1)
- Quite a bit (2)
- Somewhat (3)
- Very little (4)
- Not at all (5)

Q42 To what extent does the ${e://Field/Country2} government influence the day-to-day operations of the United States government?

- A great deal (1)
- Quite a bit (2)
- Somewhat (3)
- Very little (4)
- Not at all (5)

Page Break

Q44 How believable did you find this scenario?

- Very believable (1)
- Somewhat believable (2)
- Somewhat unbelievable (3)
- Not believable at all (4)

Page Break

Q45 Do you have any comments for the researchers conducting this study?

_____
_____
_____
_____
_____

End of Block: Block 5

Chapter 5: Conclusion

State propagandists have long capitalized on the technology of the day to wage covert, informational campaigns against foreign publics. Today, social media offers newfound scale, speed, and anonymity for such campaigns, all at low cost. As a result, it should come as no surprise that social media influence operations are occurring, or at least are discovered, more frequently.[1] Yet, International Security Studies has left the phenomenon relatively unexamined. In the three core papers of the dissertation, I made distinct contributions to further understanding of the psychological mechanisms and potential effects of foreign influence operations. Together, these papers offer a conceptualization of the phenomenon, present original data on possible effects, and test the benefits of secrecy for informational campaigns.

In this chapter, I briefly revisit the contributions of the core dissertation papers, explore their implications, and offer avenues for future research on influence operations.

Contributions & Implications

My first dissertation paper, "Foreign Influence Operations and International Security Studies" lays the foundation for a budding research program on foreign influence operations in International Relations and International Security. Existing studies diverge over how to define influence operations and how influence operations relate to other tools of statecraft.[2] I defined influence operations as "covert, informational efforts intended to affect attitudes or opinions in a manner favorable to the perpetrator." I showed different 'covert' strategies of influence operations that political actors use in the cyber age and have used in other periods. These include covert identity, covert sourcing, covert coordination, and covert sponsorship. I also applied theories from the kinetic covert action literature to suggest that states may opt for secrecy for two categories of reasons: (1) secrecy for operational effectiveness, and (2) secrecy to avoid political

[1] M.A., Renée DiResta, Josh A. Goldstein, and Shelby Grossman. "Middle East Influence Operations: Observations Across Social Media Takedowns." Paper prepared for the Project on Middle East Political Science's 2021 Conference, "Digital Activism and Authoritarian Adaptation in the Middle East."
[2] Pamment, James and Alicia Wanless. 2019. "How Do You Define a Problem Like Influence?" *Journal of Information Warfare* 18(3): 1-14.

repercussions. These potential rationales for secrecy motivate the empirical tests in subsequent dissertation papers.[3]

In addition to providing a conceptualization of influence operations, my first paper argued that influence operations are prevalent and are likely to remain so for three reasons. First, they are cost-effective. Social media influence operations are cheap, making the barrier to entry low. The low cost of influence operations helps to explain the diversity of actors that have been implicated in recent operations; influence operations are used by great powers and small ones, and by a variety of non-state actors as well.[4] Second, foreign influence operations are difficult to deter. Democratic societies offer many potential avenues of influence for a perpetrator, including diverse broadcast and social media channels. Deterrence-by-denial, or decreasing the likelihood of an operation's success, is challenging because it requires building defenses or resiliency along many channels that are not typically controlled by the government. Deterrence-by-punishment, or decreasing the likelihood of an operation by threatening costly punishment, is challenging too, because governments want to avoid retaliating against influence operations in a way that could lead to escalation. Third, propagandists are prone to inflate the impact of influence operations, even if single operations are ineffective. This likely creates a positive feedback loop, where actors who use influence operations will be tempted to do so again in the future. Together, these rationales both help explain the contemporary prevalence of social media influence operations and their likely staying power.

My second paper, "Follow the Experts or Follow the Crowd? Experimental Evidence on Public Attitudes Towards a Cyber-Attack," tests whether a common mechanism that foreign actors deploy in influence operations—influencing target attitudes via peer cues—is a viable strategy. In other words, can foreign actors benefit from secrecy for operational effectiveness

---

[3] Carnegie, Allison. 2021. "Secrecy in International Relations and Foreign Policy." *Annual Review of Political Science* 24: 213-233; Carson, Austin. 2018. *Secret Wars: Covert Conflict in International Politics*. Princeton, New Jersey: Princeton University Press; Poznansky, Michael. 2020. "Revisiting plausible deniability." *Journal of Strategic Studies*: 1-23.

[4] For Facebook's attribution of operations to a diverse set of actors, see: Gleicher, Nathaniel, Margarita Franklin, David Agranovich, Ben Nimmo, Olga Belogolova, and Mike Torrey. 2021. "The State of Influence Operations 2017-2020" *Facebook Report*. Accessed: https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf; For report documents different types of perpetrators of influence operations, see: Bradshaw, Samantha and Philip N. Howard. 2017. "Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation." *Oxford Internet Institute Report;* Bradshaw, Samantha and Philip N. Howard. 2019. "The Global Disinformation Disorder: 2019 Global Inventory of Organised Social Media Manipulation." *Oxford Internet Institute Report*.

when they falsify peer or public cues? Existing literature in International Relations on public opinion formation about foreign policy posits that citizens take cues from their favored elites.[5] The conventional wisdom leaves little room for peer or public cues, which makes the tendency of propagandists to pose as ordinary citizens puzzling.[6] Leveraging nationally representative, cross-country survey experiments in the United States and United Kingdom, I test the relative effect of expert and public cues on respondents' support for the use of force in response to a hypothetical cyber-attack. Expert and public cues both affect respondent preferences, and expert cue effects do not significantly outweigh public ones. This does not show conclusively that influence operations where actors pose as members of the target state have large impacts, but it suggests that the "Sikkinos Strategy," according to which actors conceal their true identity to persuade a target, may be effective. It also significantly challenges the conventional International Relations wisdom, highlighting the need for scholars to pay more attention to the diverse cues that reach citizens in the cyber age.

My third paper, "Plausible Deniability and Outsourced Influence Operations," examines a rising trend: the outsourcing of influence operations to private marketing and PR firms. The few studies that address outsourcing suggest that political actors outsource disinformation campaigns, at least in part, for the benefits of plausible deniability.[7] However, this assumption has yet to be tested.[8] We do not know how states benefit from plausible or implausible deniability, and how citizens in the target state view foreign operations that are sponsored or conducted by a foreign government compared to those in which the government's role is not known. I theorize different types of uncertainty about foreign influence operations, given that information about operations varies widely.[9] When it comes to foreign influence operations, there may be uncertainty about (a) whether an operation has occurred at all, (b) whether an operation originates from domestic or foreign sources, and, if it originates abroad, (c) which particular foreign actor is behind the

[5] E.g. Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press; Berinsky, Adam J. 2007. "Assuming the Costs of War: Events, Elites, and American Public Support for Military Conflict." *The Journal of Politics* 69(4): 975–97.
[6] For an exception, see: Kertzer, Joshua D. and Thomas Zeitzoff. 2017. "A Bottom-Up Theory of Public Opinion about Foreign Policy." *American Journal of Political Science* 61(3): 543–58.
[7] DiResta, Renée, Shelby Grossman, and Alexandra Siegel. "In-House vs. Outsourced Trolls: Digital Mercenaries Shape State Influence Strategies." *Working Paper.*
[8] The closest test is a survey experiment that looks at the use of proxies in the kinetic space. See: Williamson, Scott. "Proxies and Plausible Deniability" *Working Paper.*
[9] Lin, Herbert. 2016. "Attribution of Malicious Cyber Incidents: From Soup to Nuts." *Journal of International Affairs.* 70(1): 75-137.

campaign. I then deployed a survey experiment that varies whether different foreign governments are implicated in an influence operation targeting the 2024 U.S. presidential election campaign in different ways. I find that when governments are not implicated in operating or paying for an influence operation that originates in their country, they can avoid blowback; and, surprisingly, these plausibly deniable operations are no less effective at stirring paranoia or giving the impression of omnipotence than when the government's involvement is known. These results suggest that the likely calculation of political actors outsourcing influence operations to third-party firms is correct. Political actors can reap the benefits of plausible deniability and do not forego the supposed benefits of getting caught.

Implications

As Graham Allison often says, "Diagnosis must precede prescription."[10] My dissertation was primarily about diagnosing the problem, focusing on how influence operations work and what potential effects they may have. Many public organizations and initiatives are devoted to prescription and implementation—tracking and exposing influence operations, providing tools that assess authenticity of accounts, fact checking, and more.[11] A proper treatment of these potential defense mechanisms is outside the scope of this dissertation. Nevertheless, I offer a few brief reflections on the implications of my two papers that present data from survey experiments for policymakers and the broader field of International Relations.

My second paper found that, contrary to expectations developed in prior work, elite cues do not have a significantly stronger impact than public cues on American and British citizens' foreign policy preferences in response to a hypothetical cyber-attack. If propagandists who pretend to be citizens in the target state can influence citizens' preferences, democratic states face the challenge of how to prevent such influence without infringing on privacy or endangering activities. Social media platforms could, for example, require identity verification to obtain an account. However, this has real risks. Governments could pressure social media companies to

[10] For example, Opening Statement by Dr. Graham T. Allison Before the United States Senate Foreign Relations Committee on East Asia, The Pacific, and Cybersecurity Policy. "American Leadership in the Asia Pacific, Part 4: The View from Beijing." November 14, 2017. Accessed: https://www.foreign.senate.gov/imo/media/doc/111417_Allison_Testimony.pdf

[11] Smith, Victoria. 2020. "Mapping Worldwide Initiatives to Counter Influence Operations." *Carnegie Endowment for International Peace*. Accessed: https://carnegieendowment.org/2020/12/14/mapping-worldwide-initiatives-to-counter-influence-operations-pub-83435

provide access to this information. Hackers could obtain troves of personal data. Smaller steps are possible too; in co-authored work I have suggested, for example, that Twitter provide a transparency feature allowing users to check whether an account switches its handle.[12] This would prevent actors from building up large followings using follow-for-follow behavior, and then, after amassing a large following, transitioning to political content. Further work is necessary to reimagine possibilities for preventing the creation of agents-of-influence without putting authentic users at risk.

My third paper investigated outsourced disinformation. My findings suggest outsourcing may be an effective strategy: governments that are not implicated in conducting or sponsoring an operation benefit from plausible deniability and do not forego the supposed benefits of getting caught: an air of omnipotence. As the cost-benefit calculation of outsourcing is hence positive for states involved, the counter-influence operations community should explore different ways to increase the costs for private firms that wage deceptive campaigns. Grossman and Ramali highlight that a marketing firm responsible for running a disinformation campaign on Twitter also represents corporate clients such as Dunkin' Donuts. The authors write, "Advocacy groups could pressure companies like Dunkin' Donuts to avoid working with digital marketing firms that have been linked to disinformation operations."[13] As I describe in the future research section below, additional work is needed to determine the effect of naming and shaming strategies.

Future Research

As outlined in Chapter 1, this dissertation does not seek to establish the last word on the effects of foreign influence operations. Rather my first paper, in particular, seeks to serve as the foundation for a growing research program—hopefully in both International Security Studies and Political Science. In this section, I offer a number of potential avenues for research that builds on this dissertation.

[12] Goldstein, Josh A. and Shelby Grossman. "How Disinformation Evolved in 2020." *Brookings TechStream.* January 4, 2021. Accessed: https://www.brookings.edu/techstream/how-disinformation-evolved-in-2020/
[13] Grossman, Shelby and Khadeja Ramali. 2020. "Outsourcing Disinformation." *Lawfare*. Accessed: https://www.lawfareblog.com/outsourcing-disinformation

1. Domestic versus Foreign Influence Operations

My dissertation primarily focuses on the mechanisms and potential effects of *foreign* influence operations. However, social media platforms increasingly find that coordinated inauthentic behavior is run from domestic, not foreign, sources.[14] In the last year alone, multiple militaries (or people associated with militaries) have been caught running influence operations targeting their own people.[15] Domestic political actors have been caught using fake accounts for intra-party politics.[16] Limited research has investigated how domestic versus foreign influence operations differ, and whether public expectations about platform policy vary based on the identity of the source of information.

A basic descriptive agenda could begin by studying how the tactics used in domestic operations are similar to or different from those used in foreign operations. One potential limitation on the effectiveness of foreign operations is the perpetrator's ability to create persuasive messaging. This requires knowledge of nuances in the target society and the ability to effectively communicate as native speakers, lest risk getting caught. Domestic propagandists have a comparative advantage: they likely have a nuanced understanding of their own culture, they often speak the language of their target, and thus, ceteris paribus, may be able to develop more persuasive messaging. A research project could begin by looking at Twitter's database of online influence operations content for similarities and differences between domestic and foreign campaigns. For example, do domestic operations receive more engagement by authentic users than operations run by foreigners? Do domestic operators do a better job evading detection and thus operate for longer periods of time on the platform before they are removed? Descriptive research would help establish the relative importance of the domestic and foreign challenges. A second research project might consider citizens' expectations of information transparency. Do

[14] Gleicher, Nathaniel, Margarita Franklin, David Agranovich, Ben Nimmo, Olga Belogolova, and Mike Torrey. 2021. "The State of Influence Operations 2017-2020" *Facebook Report*. Accessed: https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf

[15] For two examples, see: Goldstein, Josh A., Aim Sinpeng, Daniel Bush, Ross Ewald, and Jennifer John. 2020. "Cheerleading Without Fans: A Low-Impact Domestic Information Operation by the Royal Thai Army." *Stanford Internet Observatory Report*; Grossman, Shelby, Elena Cryst, Renée DiResta, K.H., Carly Miller, Rajeev Sharma, Chase Small, and Julia Thompson. 2021. "My Heart Loves the Army: An Investigation into a Jordanian Disinformation Campaign on Facebook, TikTok, and Twitter." *Stanford Internet Observatory Report*.

[16] For example, see: Grossman, Shelby, Gil Baram, Josh A. Goldstein, and Carly Miller. 2021. "Staying Current; an Investigation into a Now-Suspended Facebook Network Promoting the Palestinian Democratic Reform Current." *Stanford Internet Observatory Report*.

citizens disapprove equally of covert operations from domestic versus foreign sources? If not, what explains the variation in user expectation around content from these two types of sources?

2. Diagnosing the Influence Industry

My second paper established that non-elite cues can impact people's foreign policy preferences. My third paper established that political actors can benefit from plausible deniability when waging influence operations if they are not implicated by an attribution. Additional research could build on these findings by studying disinformation-for-hire or investigating a broader set of elite and non-elite cues in the social media age.

One research program could build on existing studies and map known private firms that offer deceptive social media services to political actors. Political scientists might explore variation in the location of such firms. Possible explanatory factors include the host state's media environment, job prospects, and digital literacy rates. Scholars could also study the effect of naming and shaming firms that are implicated in running influence operations.
Do such firms reestablish themselves under different names? At what rate do they manage to continue their deceptive activities on the same (or different) social media platforms? Precedent exists for fruitful collaboration between investigative journalists and political scientists in the counter-disinformation field.[17] Such cooperation may be particularly promising for investigating the effects of attribution on disinformants-for-hire.

Scholars could also contribute to cue-taking literature by testing the impact that a wider variety of elite and non-elite cues have on citizens' foreign policy preferences. In my second paper, I used a weak form of a peer cue—public opinion writ large—to intentionally create a hard case for the bottom-up theory of opinion formation and an easy case for the top-down conventional wisdom.[18] Citizens are more likely to be influenced by those they deem as credible and trustworthy.[19] Future research could look at different types of influential cues in the social media age that share political content or content about foreign affairs. This may help explain why

[17] Ward, Clarissa, Katie Polglase, Sebastian Shukla, Gianluca Mezzofiore, and Tim Lister. "Russian election meddling is back – via Ghana and Nigeria – in your feeds." *CNN*. April 11, 2020. Accessed: https://www.cnn.com/2020/03/12/world/russia-ghana-troll-farms-2020-ward/index.html
[18] Kertzer, Joshua D. and Thomas Zeitzoff. 2017. "A Bottom-Up Theory of Public Opinion about Foreign Policy." *American Journal of Political Science* 61(3): 543–58.
[19] Hovland, Carl I. and Walter Weiss. 1951. "The Influence of Source Credibility on Communication Effectiveness." *Public Opinion Quarterly* 15: 635-650.

governments co-opt or attempt to silence particular types of social media influencers.[20] The implications here extend far beyond covert informational campaigns; they are crucial for reimagining theories of public opinion formation in International Relations for the cyber age.

3. Influence Operations and Emerging Technology

As I described in Chapter 1, propagandists capitalize on the technology of the day for their influence operations. Since the early 2010s, social media has offered propagandists an established tool to reach foreign publics, leading to the rise of cyber-enabled influence operations. As mentioned in my first paper, propagandists have already innovated in response to defender strategies within the domain of social media. For instance, propagandists have used AI-generated profile pictures to avoid reverse image searching.[21] While existing research has begun to explore how emerging AI technologies may be used for propaganda purposes, there are opportunities for additional contributions.

Multiple reports establish that AI text-generating technologies can write articles and tweet-length content. For example, Kreps, McCain, and Brundage ran survey experiments in which they presented some respondents with text written by humans for real news outlets and others with content written by OpenAI's GPT-2. The authors found that "individuals are largely incapable of distinguishing between AI- and human-generated text."[22] Researchers from Georgetown's Center for Security and Emerging Technology experimented with OpenAI's GPT-3 and found that it could create slanted content and invent conspiracies.[23] Additional research could model how such language-generating AI technologies may be used by foreign propagandists. Other research could test how deepfake videos may be used in crisis scenarios and the extent to which ordinary users are suspicious of political information from social media accounts that use AI-generated profile pictures.

---

[20] For a recent example, see: McGee, Luke. "The brazen arrest of a Belarusian activist has terrified dissidents all over the world." CNN. May 30, 2021. Accessed: https://www.cnn.com/2021/05/29/europe/belarus-arrest-ryanair-dissidents-unsafe-intl-cmd/index.html

[21] Goldstein and Grossman 2021.

[22] Kreps, Sarah, R. Miles McCain, and Miles Brundage. 2020. "All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation." *Journal of Experimental Political Science* 1-14.

[23] Buchanan, Ben, Andrew Lohn, Micah Musser, and Katerina Sedova. 2021. "Truth, Lies, and Automation: How Language Models Could Change Disinformation." *Center for Security and Emerging Technology Report*. Accessed: https://cset.georgetown.edu/publication/truth-lies-and-automation/

4. Preventing Perception Hacking

After 2016, Americans were fearful of the effect of foreign interference in the 2020 election. According to reporting from October 2020, 59% of Americans said that they were "extremely" or "somewhat' concerned about potential election interference by a foreign government in the 2020 election.[24] Existing work in Political Science shows that even small forms of foreign interference erode trust in an election under hypothetical conditions.[25] In my third paper, I found that a substantial percentage of survey respondents believed that thousands of fake social media accounts had an effect on the 2024 election results—even though they were not given any information about spread or engagement (e.g. how many real Americans saw content from these accounts, how many people shared the content). Additional research might explore the potential benefits and drawbacks of normalizing citizens to the fact that fake accounts exist on social media platforms. Even if social media companies improve enforcement, networks of accounts will not become prohibitively costly for well-resourced political actors. It is inevitable that some networks will continue to arise (and get caught).

On the one hand, such normalization could mitigate the potential for foreign actors to engage in "perception hacking," the term that Facebook uses for attempts to "play on our collective expectation of wide-spread interference to create the perception that they're more impactful than they in fact are."[26] Perception hacking is particularly dangerous around elections because if foreign actors can convince citizens in the target state that the election was not domestically determined, they can undermine faith in a core pillar of democracy. Normalizing citizens to the presence of fake accounts, however, may have costs. It could further erode trust in the information environment and give citizens reason to write off political stances they disagree with as the product of foreign manipulation. Future research could rely on real public messaging by civil society actors or social media companies about the presence and/or prevalence of such fake accounts. A survey could assess the effects of such information on both perceptions of a subsequent announcement of an interference attempt and on broader outcomes like trust in news.

---

[24] Miller, Maggie. "Majority of Americans concerned about potential foreign interference: poll." *The Hill*. October 2, 2020. Accessed: https://thehill.com/policy/cybersecurity/519411-majority-of-americans-concerned-about-potential-foreign-election

[25] Tomz, Michael and Jessica L. P. Weeks. 2020. "Public Opinion and Foreign Electoral Intervention." *American Political Science Review* 114(3): 856-873.

[26] Gleicher, Nathaniel. "Removing Coordinated Inauthentic Behavior." *Facebook Blog*. October 27, 2020. Accessed: https://about.fb.com/news/2020/10/removing-coordinated-inauthentic-behavior-mexico-iran-myanmar/

This would help weigh the benefits of a broader informational campaign for creating resiliency against the potential costs of undermining trust or increasing paranoia.

5. Normative Theory and Going on the Offense

In my second and third papers, I conduct original survey experiments on American and British citizens to investigate public opinion formation. In much of the framing, I also talk about the threat of influence operations to democratic societies specifically. However, we know that the United States and the United Kingdom are not merely the targets of influence operations, but have also waged influence operations themselves. As the prevalence of social media influence operations grows, people have begun, and are likely to continue, calling on Western governments to use active measures more frequently. Research should explore the potential costs and benefits of action before such calls become more widespread.

There are both empirical and normative methods to pursue this question. Empirically, the utility of influence operations is a function of, among other things, how effective an operation is, the repercussions if the operation is exposed, and the likelihood of exposure. Survey experiments could examine the conditions under which publics support or oppose their own government using active measures abroad. Studies could also look at how attribution of an influence operation affects a country's reputation by citizens in third-party countries. Reputational backlash among domestic citizens, members of the target state, or members of third-party countries are different potential disincentives for democratic states. From a normative perspective, scholars could examine the trade-offs between norms of transparency and the potential benefits of an operation.[27] Who is a legitimate target for disinformation campaigns or active measures? Is there ever a legitimate target?

[27] Myrick, Rachel. 2020. "Why So Secretive? Unpacking Public Attitudes toward Secrecy and Success in US Foreign Policy." *Journal of Politics* 82(3): 828-843.

Bibliography:

Agadjanian, Alexander. 2021. "When Do Partisans Stop Following the Leader?" *Political Communication* 38(4): 351-369.

Agadjanian, Alexander and Yusaku Horiuchi. 2020. "Has Trump Damaged the U.S. Image Abroad? Decomposing the Effects of Policy Messages on Foreign Public Opinion." *Political Behavior* 42(2): 581–602.

Asal, Victor, Luis De La Calle, Michael Findley, and Joseph Young. 2012. "Killing Civilians or Holding Territory? How to Think about Terrorism." *International Studies Review* 14: 475-497.

Allen, Susan Hannah. 2008. "The Domestic Political Costs of Economic Sanctions." *The Journal of Conflict Resolution* 52(6): 916–944.

Allison, Graham T. 1971. *Essence of Decision: Explaining the Cuban Missile Crisis*. Boston: Little, Brown and Company.

Allison, Graham T. and Morton H. Halperin. 1972. "Bureaucratic Politics: A Paradigm and Some Policy Implications." *World Politics*. 24: 40-79.

Almond, Gabriel A. 1950. *The American People and Foreign Policy*. 1st ed. New York: Harcourt.

Almond, Gabriel A. 1956. "Public Opinion and National Security Policy." *Public Opinion Quarterly* 20(2): 371-378.

Ames, Daniel R. and Lara K. Kammrath. 2004. "Mind-Reading and Metacognition: Narcissism, Not Actual Competence, Predicts Self-Estimated Ability. *Journal of Nonverbal Behavior* 28(3): 187-209.

Andrew Christopher and David Dilks. 1984. *The Missing Dimension: Governments and Intelligence Communities in the Twentieth Century*. Basingstoke: Macmillan.

Andrew, Christopher and Vasili Mitrokhin. 2001. *The Sword and the Shield: The Mitrokhin Archive and the Secret History of the KGB.* New York: Basic Books.

Andrew, Christopher and Oleg Gordievsky. 1990. *KGB The Inside Story.* New York: HarperCollins Publishers.

Ansolabehere, Stephen and Shanto Iyengar. 1994. "Of Horseshoes and Horse Races: Experimental Studies of the Impact of Poll Results on Electoral Behavior." *Political Communication* 11(4): 413–30.

Anspach, Nicolas M. 2017. “The New Personal Influence: How Our Facebook Friends Influence the News We Read.” *Political Communication* 34: 590-606.

Arceneaux, Kevin. 2008. “Can Partisan Cues Diminish Democratic Accountability?” *Political Behavior* 30(2): 139–60.

Asch, Solomon E. 1951. “Effects of Group Pressure upon the Modification and Distortion of Judgments.” In *Groups, Leadership and Men; Research in Human Relations*, edited by Harold Steere Guetzkow, 177–90. Pittsburgh, PA: Carnegie Press.

Associated Press. “AP Fact Check: Coca-Cola not recalling Dasani water.” *Associated Press*. January 24, 2017. Accessed: https://apnews.com/article/ecfb398fa555456d99fe40e0a05dd1da

Associated Press. “Three in four Americans are worried about foreign interference in the presidential election.” *Associated Press.* October 2, 2020. Accessed: https://www.marketwatch.com/story/most-americans-are-worried-about-foreign-interference-in-the-election-01601648756

Badawy, Adam, Kristina Lerman, and Emilio Ferrara. 2019. “Who Falls for Online Political Manipulation?” In *Companion Proceedings of the 2019 World Wide Web Conference (WWW’19 Companion)*: 162-168.

Bail, Christopher A., Brian Guay, Emily Maloney, Aidan Combs, D. Sunshine Hillygus, Friedolin Merhout, Deen Freelon, and Alexander Volfovsky. 2020. “Assessing the Russian Internet Research Agency’s impact on the political attitudes and behaviors of American Twitter users in late 2017.” *PNAS* 117(1): 243-250.

Barberá, Pablo, Andreu Casas, Jonathan Nagler, Patrick J. Egan, Richard Bonneau, John T. Jost, and Joshua A. Tucker. 2019. "Who leads? Who follows? Measuring issue attention and agenda setting by legislators and the mass public using social media data." *American Political Science Review* 113(4): 883-901.

Barberá, Pablo and Thomas Zeitzoff. 2018. “The New Public Address System: Why Do World Leaders Adopt Social Media?” *International Studies Quarterly* 62(1): 121-130.

Barnes, Julian E. “State Dept. Traces Russian Disinformation Links.” *The New York Times*. August 5, 2020. Accessed: https://www.nytimes.com/2020/08/05/us/politics/state-department-russian-disinformation.html

Barrett, Paul M. 2019. “Disinformation and the 2020 Election: How the Social Media Industry Should Prepare” *NYU Stern Center for Business and Human Rights Report*. Accessed: https://issuu.com/nyusterncenterforbusinessandhumanri/docs/nyu_election_2020_report/1

Barrie, Christopher and Alexandra A. Siegel. 2021. “Kingdom of Trolls? Influence Operations in the Saudi Twittersphere.” *Journal of Quantitative Description: Digital Media* 1: 1-41.

Bastos, Marco and Johan Farkas. 2019. “‘Donald Trump Is My President!’: The Internet Research Agency Propaganda Machine.” *Social Media + Society*: 1-13.

Bateman, Jon, Elonnai Hickock, Laura Courchesne, Isra Thange, and Jacob N. Shapiro. 2021. “Measuring the Effects of Influence Operations: Key Findings and Gaps From Empirical Research.” *Carnegie Endowment for International Peace*. Accessed: https://carnegieendowment.org/2021/06/28/measuring-effects-of-influence-operations-key-findings-and-gaps-from-empirical-research-pub-84824

Baum, Matthew A. 2004. “How Public Opinion Constrains the Use of Force: The Case of Operation Restore Hope.” *Presidential Studies Quarterly* 34(2): 187-226.

Baum, Matthew A. and Philip B. K. Potter. 2019. “Media, Public Opinion, and Foreign Policy in the Age of Social Media.” *The Journal of Politics* 81(2): 747-756.

Baum, Matthew A. and Tim Groeling. 2009. “Shot by the Messenger: Partisan Cues and Public Opinion Regarding National Security and War.” *Political Behavior* 31: 157-186.

Baum, Matthew A. and Tim Groeling. 2010. “Reality Asserts Itself: Public Opinion on Iraq and the Elasticity of Reality.” *International Organization* 64(3): 443–79.

Beckley, Michael. 2018. “The Power of Nations: Measuring What Matters.” *International Security* 43(2): 7-44.

Belfer Center, “U.S. Not Prepared for Major Cyber Attack.” *NBC Today Show*, December 26, 2018. Accessed: https://www.belfercenter.org/publication/eric-rosenbach-nbc-us-not-prepared-major-cyber-attack.

Benkler, Yochai. 2020. “The Danger of Overstating the Impact of Information Operations.” *Lawfare*. Accessed: https://www.lawfareblog.com/danger-overstating-impact-information-operations.

Benkler, Yochai, Robert Faris, and Hal Roberts. 2018. *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.

Berinsky, Adam J. 2007. “Assuming the Costs of War: Events, Elites, and American Public Support for Military Conflict.” *The Journal of Politics* 69(4): 975–97.

Berinsky, Adam J. 2009. *In Time of War: Understanding American Public Opinion from World War II to Iraq*. Chicago Studies in American Politics. Chicago: University of Chicago Press

Bisgaard, Martin. 2019. “How Getting the Facts Right Can Fuel Partisan-Motivated Reasoning.” *American Journal of Political Science* 63(4): 824-839.

Block, David and Richard A. Harper. 1991. “David Block and Richard A. Harper, “Overconfidence in Estimation: Testing the Anchoring-and-Adjustment Hypothesis.” *Organizational Behavior and Human Decision Processes* 49: 188-207.

Bode, Leticia and Emily K. Vraga. 2021. “Correction Experiences on Social Media During COVID-19.” *Social Media + Society*: 1-11.

Boettcher III, William A. and Michael D. Cobb. 2006. “Echoes of Vietnam? Casualty Framing and Public Perceptions of Success and Failure in Iraq.” *Journal of Conflict Resolution* 50(6): 831-854.

Bolsen, Toby, James N. Druckman, and Fay Lomax Cook. 2014. “The Influence of Partisan Motivated Reasoning on Public Opinion.” *Political Behavior* 36(2): 235-262

Borghard, Erica. 2019. “What A U.S. Operation Against Russian Trolls Predicts About Escalation in Cyberspace.” *War on the Rocks*. Accessed: https://warontherocks.com/2019/03/what-a-u-s-operation-against-russian-trolls-predicts-about-escalation-in-cyberspace/

Boyd, Robert, and Peter J. Richerson. 1985. *Culture and the Evolutionary Process*. Chicago: University of Chicago Press

Boyd, Robert, Peter J. Richerson, and Joseph Henrich. 2011. “The Cultural Niche: Why Social Learning Is Essential for Human Adaptation.” *Proceedings of the National Academy of Sciences of the United States of America* 108: 10918–10925.

Bradley, James V. 1981. “Overconfidence in Ignorant Experts.” *Bulletin of the Psychonomic Society* 17(2): 82–84.

Bradshaw, Samantha and Philip N. Howard. 2017. “Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation.” *Oxford Internet Institute Report.*

Bradshaw, Samantha and Philip N. Howard. 2019. “The Global Disinformation Disorder: 2019 Global Inventory of Organised Social Media Manipulation.” *Oxford Internet Institute Report*.

Busby, Joshua W. and Jonathan Monten. 2012. “Republican Elites and Foreign Policy Attitudes.” *Political Science Quarterly* 127(1): 105–42.

Busby, Joshua, Craig Kafura, Jonathan Monten, and Jordan Tama. 2019. “Multilateralism and the Use of Force: Experimental Evidence on the Views of Foreign Policy Elites.” *Foreign Policy Analysis* 16(1): 118-129.

Buzan, Barry and Lene Hansen. 2009. “Defining International Security Studies” in *The Evolution of International Security Studies*. Cambridge University Press.

Carmines, Edward G. and James A. Stimson. 1980. “The Two Faces of Issue Voting.” *American Political Science Review* 74(1): 78–91.

Carmines, Edward G. and James H. Kuklinski. 1990. “Incentives, opportunities, and the logic of public opinion in American political representation.” In J.A. Ferejohn and J.H. Kuklinski (eds.), *Information and Democratic Processes*, pp. 240–268. Urbana: University of Illinois Press.

Carnegie, Allison. 2021. “Secrecy in International Relations and Foreign Policy.” *Annual Review of Political Science* 24: 213-233.

Carrier-Sabourin, Krystel. 2011. “Measuring Effects and Success in Influence Operations: Challenges, Limitations, and Opportunities,” Contract Report, Defence R&D Canada. Accessed: https://apps.dtic.mil/dtic/tr/fulltext/u2/a568381.pdf.

Carson, Austin. 2016. “Facing Off and Saving Face: Covert Intervention and Escalation Management in the Korean War.” *International Organization* 70(1): 103-131.

Carson, Austin. 2018. *Secret Wars: Covert Conflict in International Politics*. Princeton, New Jersey: Princeton University Press.

Carson, Austin and Keren Yarhi-Milo. 2017. “Covert communication: The Intelligibility and Credibility of Signaling in Secret.” *Security Studies* 26(1): 124-156.

Caspary, William R. 1970. “The “Mood Theory”: A Study of Public Opinion and Foreign Policy.” *American Political Science Review* 64(2): 536-547.

Cederberg, Gabriel, Jordan D’Amato, Corinna Fehst, Simon Jones, Kunal Kothari, Aleksandra Milcheva, and Irene Solaiman. 2019. “National Counter-Information Operations Strategy.” *Belfer Center Report.* Accessed: https://www.belfercenter.org/publication/national-counter-information-operations-strategy

Chen, Adrian. “The Real Paranoia-Inducing Purpose of Russian Hacks.” *The New Yorker*. July 27, 2016. Accessed: https://www.newyorker.com/news/news-desk/the-real-paranoia-inducing-purpose-of-russian-hacks

Chu, Jonathan and Stefan Recchia. 2020. “Does Public Opinion affect the Preferences of Foreign Policy Leaders? Experimental Evidence from the UK Parliament.” *Working Paper.*

Chudek, Maciej, Sarah Heller, Susan Birch, and Joseph Henrich. 2012. "Prestige-biased cultural learning: Bystander's differential attention to potential models influences children's learning." *Evolution and Human Behavior* 33(1): 46-56.

Cialdini, Robert B. 2009. *Influence: Science and Practice*. 5th ed. Boston: Pearson/Allyn and Bacon.

Cialdini, Robert B. and Noah J. Goldstein. 2004. "Social Influence: Compliance and Conformity." *Annual Review of Psychology* 55: 591-621.

Clarke, Killian and Korhan Kocak. 2020. "Launching Revolution: Social Media and the Egyptian Uprising's First Movers." *British Journal of Political Science* 50: 1025-1045.

Cohen, Eliot A. 1996. "A Revolution in Warfare." *Foreign Affairs* 75(2): 37-54.

Converse, Philip E. 2006. "The nature of belief systems in mass publics" (1964), Critical Review. 18:1-3, 1-74.

Coppock, Alexander and Oliver A. McClellan. 2019. "Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents." *Research and Politics* 1-14.

Cormac, Rory. 2019. "Techniques of Covert Propaganda: The British Approach in the Mid-1960s." *Intelligence and National Security* 34(7): 1064-1069.

Cormac, Rory, Calder Walton, and Damien Van Puyvelde. 2021. "What constitutes successful covert action? Evaluating unacknowledged interventionism in foreign affairs." *Review of International Studies* 1-18.

Cormac, Rory and Richard J. Aldrich. 2018. "Grey is the new black: covert action and implausible deniability" *International Affairs* 94(3): 477-494.

Corstange, Daniel and Nikolay Marinov. 2012. "Taking Sides in Other People's Elections: The Polarizing Effect of Foreign Intervention." *American Journal of Political Science* 56(3): 655-670.

Crossman, R. H. S. 1952. "Psychological Warfare." *Royal United Services Institution Journal* 98(592): 528-539.

Crutchfield, Richard S. 1955. "Conformity and Character." *American Psychologist* 10(5): 191-198: 191–98.

Cryst, Elena, Esteban Ponce de León, and Daniel Suárez Perez. 2020. "Bolivarian Factions: Facebook Takes Down Inauthentic Assets." *Stanford Internet Observatory Report*. Accessed: https://fsi.stanford.edu/publication/bolivarian-factions-facebook-takes-down-inauthentic-assets

Dahl, Robert A. 1957. "The Concept of Power." *Behavioral Science* 2(3): 201-215.

Deutsch, Morton, and Harold B. Gerard. 1955. "A Study of Normative and Informational Social Influences upon Individual Judgment." *The Journal of Abnormal and Social Psychology* 51(3): 629–36.

Diamond, Larry. 2010. "Liberation Technology." *Journal of Democracy* 21(3): 69-83.

Dietrich, Simone, Heidi Hardt, and Haley J. Swedlund. 2021. "How to make elite experiments work in International Relations." *European Journal of International Relations* 27(2): 596-621.

Dimmick, John, Yan Chen, and Zhan Li. 2004. "Competition Between the Internet and Traditional News Media: The Gratification-Opportunities Niche Dimension." *The Journal of Media Economics* 17(1): 19-33.

DiResta, Renée. 2018. "Open Hearing on Foreign Influence Operations' Use of Social Media Platforms (Third Party Expert Witnesses)." § Senate Select Committee on Intelligence. Accessed: https://www.justsecurity.org/wp-content/uploads/2019/07/Open-Hearing-on-Foreign-Influence-Operations%E2%80%99-Use-of-Social-Media-Platforms-Third-Party-Expert-Witnesses.pdf

DiResta, Renée. 2018. "Computational Propaganda: If You Make It Trend, You Make It True." *The Yale Review* 106(4): 12-29.

DiResta, Renée, Kris Shaffer, Becky Ruppel, David Sullivan, Robert Matney, Ryan Fox, Jonathan Albright, and Ben Johnson. 2018. "The Tactics & Tropes of the Internet Research Agency." *New Knowledge Report*. Accessed: https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1003&context=senatedocs

DiResta, Renée, Michael McFaul, and Alex Stamos. "Here's how Russia will attack the 2020 election. We're still not ready." *The Washington Post*. November 15, 2019. Accessed: https://www.washingtonpost.com/opinions/2019/11/15/heres-how-russia-will-attack-election-were-still-not-ready/

DiResta, Renée and Shelby Grossman. 2019. "Potemkin Pages & Personas: Assessing GRU Online Operations, 2014-2019." *Stanford Internet Observatory Report.* Accessed: https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/potemkin-pages-personas-sio-wp.pdf

DiResta, Renée, Shelby Grossman, and Alexandra Siegel. "In-House vs. Outsourced Trolls: Digital Mercenaries Shape State Influence Strategies." *Working Paper.*

DiResta, Renée, Shelby Grossman, K.H., and Carly Miller. 2019. "Analysis of Twitter Takedown of State-Backed Operation Attributed to Saudi Arabian Digital Marketing Firm Smaat." *Stanford Internet Observatory Report.*

Disinformation: A Primer in Russian Active Measures and Influence Campaigns Panel II, Hearing before the Select Committee on Intelligence of the United States Senate, 105th Cong., 1st sess., March 30, 2017, 19. Accessed: https://www.justsecurity.org/wp-content/uploads/2019/07/Disinformation-A-Primer-in-Russian-Active-Measures-and-Influence-Campaigns-Part-2-Panel-II.pdf

Downes, Alexander B., and Mary Lauren Lilley. 2010. "Overt Peace, Covert War?: Covert Intervention and the Democratic Peace." *Security Studies* 19(2): 266–306.

Downs, Anthony. 1957. "An Economic Theory of Political Action in a Democracy." *Journal of Political Economy* 65(2): 135–50.

Dragojlovic, Nick. 2015. "Listening to Outsiders: The Impact of Messenger Nationality on Transnational Persuasion in the United States." *International Studies Quarterly* 59: 73-85.

Dropp, Kyle, Joshua D. Kertzer, and Thomas Zeitzoff. 2021. "The less Americans know about Ukraine's location, the more they want U.S. to intervene." *The Monkey Cage.* Accessed: https://www.washingtonpost.com/news/monkey-cage/wp/2014/04/07/the-less-americans-know-about-ukraines-location-the-more-they-want-u-s-to-intervene/

Druckman, James N. 2001. "On the Limits of Framing Effects: Who Can Frame?" *The Journal of Politics* 63(4): 1041-1066.

Duncombe, Constance. 2017. "Twitter and transformative diplomacy: social media and Iran-US relations." *International Affairs* 93(3): 545-562.

Dvir, Rotem, Nehemia Geva, and Arnold Vedlitz. 2021. "Unpacking Public Perceptions of Terrorism: Does Type of Attack Matter?" *Studies in Conflict & Terrorism*: 1-21.

Dwoskin, Elizabeth. "Russia is still the biggest player in disinformation, Facebook says." *The Washington Post.* May 26, 2021. Accessed: https://www.washingtonpost.com/technology/2021/05/26/facebook-disinformation-russia-report/

Eckstein, Harry. "Case Study and Theory in Political Science." 2009. In *Case Study Method*, edited by Roger Gomm, Martyn Hammersley, and Peter Foster. United Kingdom: SAGE Publications Ltd.

Edelman. "2019 Edelman Trust Barometer: Global Report." Accessed: https://www.edelman.com/sites/g/files/aatuss191/files/2019-02/2019_Edelman_Trust_Barometer_Global_Report.pdf

Edwards, George C. 2003. *On Deaf Ears: The Limits of the Bully Pulpit*. New Haven: London: Yale University Press.

Eichenberg, Richard C. 2016. “Public Opinion on Foreign Policy Issues.” In *Oxford Research Encyclopedia of Politics*, by Richard C. Eichenberg. Oxford University Press.

Entman, Robert M. 2004. *Projections of Power: Framing News, Public Opinion, and U.S. Foreign Policy*. Studies in Communication, Media, and Public Opinion. Chicago: University of Chicago Press.

Farrell, Henry and Bruce . 2018. “Common-Knowledge Attacks on Democracy.” *Berkman Klein Center Research Publication No. 2018-7.*

Fearon, James D. 1994. “Domestic Political Audience Costs and the Escalation of International Disputes.” *American Political Science Review* 88(3): 577-592.

Fearon, James D. 1995. "Rationalist explanations for war." *International Organization* 49(3): 379-414.

Federal Trade Commission. “Devumi, Owner and CEO Settle FTC Charges They Sold Fake Indicators of Social Media Influence; Cosmetics Firm Sunday Riley, CEO Settle FTC Charges That Employees Posted Fake Online Reviews at CEO’s Direction.” Federal Trade Commission Press Release. October 21, 2019. Accessed: https://www.ftc.gov/news-events/press-releases/2019/10/devumi-owner-ceo-settle-ftc-charges-they-sold-fake-indicators

Ferguson, Niall. 2017. *The Square and the Tower: Networks and Power from the Freemasons to Facebook*. London: Penguin Books.

Ferraro, Matthew F. and Jason C. Chipman. “Fake news threatens our businesses, not just our politics.” *The Washington Post.* February 8, 2019. Accessed: https://www.washingtonpost.com/outlook/fake-news-threatens-our-businesses-not-just-our-politics/2019/02/08/f669b62c-2b1f-11e9-984d-9b8fba003e81story.html

Fisher, Aleksandr. 2020. “Demonizing the Enemy: The Influence of Russian State-Sponsored Media on American Audiences.” *Post-Soviet Affairs*. 36(4): 281-296.

Fly, Jamie, Laura Rosenberger, and David Salvo. 2018. “Policy Blueprint for Countering Authoritarian Interference in Democracies.” *Alliance for Securing Democracy Report.* p. 10. Accessed: https://www.gmfus.org/publications/asd-policy-blueprint-countering-authoritarian-interference-democracies.

Forcese, Craig. 2016. “Pragmatism and Principle: Intelligence Agencies and International Law.” *Virginia Law Review Online* 102: 67-84.

Frenkel, Sheera and Cecilia Kang. 2021. *An Ugly Truth: Inside Facebook’s Battle for Domination.* HarperCollins Publishers, as cited in Allen, Mike. “‘An Ugly Truth’: How Facebook discovered Russian meddling.’” *Axios*. July 9, 2021. Accessed:

https://www.axios.com/facebook-book-russia-zuckerberg-sandberg-3bcb04fb-b8e9-4e8e-83d1-9a00e1218a27.html

Freedman, Lawrence. 1989. "General Deterrence and the Balance of Power." *Review of International Studies* 15(2): 199-210.

Freelon, Deen and Tetyana Lokot. 2020. "Russian Twitter disinformation campaigns reach across the American political spectrum." *The Harvard Kennedy School Misinformation Review* 1(1): 1-9.

Frye, Timothy. 2019. "Economic Sanctions and Public Opinion: Survey Experiments From Russia." *Comparative Political Studies* 52(7): 967-994.

Gabel, Matthew, and Kenneth Scheve. 2007. "Estimating the Effect of Elite Communications on Public Opinion Using Instrumental Variables." *American Journal of Political Science* 51(4): 1013–28.

Gallagher, Helen L. and Christopher D. Frith. 2003. "Functional Imaging of 'Theory of Mind.'" *Trends in Cognitive Sciences* 7(2): 77–83.

Gartner, Scott Sigmund. 1999. *Strategic Assessment in War.* New Haven; London: Yale University Press.

Gartner, Scott Sigmund. 2008. "The Multiple Effects of Casualties on Public Support for War: An Experimental Approach." *American Political Science Review* 102(1): 95–106.

Gelpi, Christopher. 2010. "Performing on Cue? The Formation of Public Opinion Toward War." *Journal of Conflict Resolution* 54(1): 88-116.

George, Alexander L., and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. BCSIA Studies in International Security. Cambridge: MIT Press.

George, Robert K. 1987. "An Historical Investigation of Soviet Strategic Deception" Report no. 87-0955 *Maxwell AFB: Air Command and Staff College*.

Gerring, John. 1999. "What Makes a Concept Good? A Criteria Framework for Understanding Concept Formation in the Social Science." *Polity* 31(3): 357-393.

Gibbs, David N. 1995. "Secrecy and International Relations." *Journal of Peace Research* 32(2): 213–28.

Glaser, Charles L. 1997. "The Security Dilemma Revisited." *World Politics* 50(1): 171-201.

Gleicher, Nathaniel. "Removing Coordinated Inauthentic Behavior." *Facebook Blog*. October 27, 2020. Accessed: https://about.fb.com/news/2020/10/removing-coordinated-inauthentic-behavior-mexico-iran-myanmar/

Gleicher, Nathaniel, Margarita Franklin, David Agranovich, Ben Nimmo, Olga Belogolova, and Mike Torrey. 2021. "The State of Influence Operations 2017-2020" *Facebook Report*. Accessed: https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf

Godson, Roy and Richard Schultz. 1985. "Soviet active measures: Distinctions and definitions." *Defense Analysis* 1(2): 101-110.

Golby, James, Peter Feaver, and Kyle Dropp. 2018. "Elite Military Cues and Public Opinion About the Use of Military Force." *Armed Forces & Society* 44(1): 44-71.

Goldman, Emily O. 2020. "From Reaction to Action: Adopting a Competitive Posture in Cyber Diplomacy." *Texas National Security Review* 3(4): 84-101.

Goldsmith, Jack and Stuart Russell. 2018. "Strengths Become Vulnerabilities: How a Digital World Disadvantages the United States In Its International Relations." *Hoover Institution Paper*. Aegis Series Paper No. 1806. Accessed: https://www.hoover.org/sites/default/files/research/docs/381100534-strengths-become-vulnerabilities.pdf

Goldstein, Josh A., Aim Sinpeng, Daniel Bush, Ross Ewald, and Jennifer John. 2020. "Cheerleading Without Fans: A Low-Impact Domestic Information Operation by the Royal Thai Army." *Stanford Internet Observatory Report*.

Goldstein, Josh A. and Lindsey Hundley. 2021. "Viva la #Resistance? Now-Suspended Twitter Accounts from Iran Claim #Resistance Personas in US Politics and Middle Eastern Affairs." *Stanford Internet Observatory Report*.

Goldstein, Josh A. and Renée DiResta. "Foreign Influence Operations and the 2020 Election: Framing the Debate." *Lawfare*. October 23, 2020. Accessed: https://www.lawfareblog.com/foreign-influence-operations-and-2020-election-framing-debate.

Goldstein, Josh A. and Shelby Grossman. "How Disinformation Evolved in 2020." *Brookings TechStream.* January 4, 2021. Accessed: https://www.brookings.edu/techstream/how-disinformation-evolved-in-2020/

Goldsmith, Benjamin E. and Yusaku Horiuchi. 2011. "Spinning the Globe? U.S. Public Diplomacy and Foreign Public Opinion." *The Journal of Politics* 71(3): 863-875

Goldsmith, Benjamin E. and Yusaki Horiuchi. 2012. "In Search of Soft Power: Does Foreign Public Opinion Matter for US Foreign Policy?" *World Politics* 64(3): 555-585.

Goldsmith, Benjamin E. and Horiuchi, Yusaku. "Does Russian Election Interference Damage Support for U.S. Alliances? The Case of Japan" *Working Paper*.

Golovchenko, Yevgeniy, Cody Buntain, Gregory Eady, Megan A. Brown, and Joshua A. Tucker. 2020. "Cross-Platform State Propaganda: Russian Trolls on Twitter and YouTube during the 2016 U.S. Election." *The International Journal of Press/Politics* 25(3): 357-389.

Gomez, Miguel Alberto 2020. "Cyber-Enabled Information Warfare and Influence Operations." *Information Warfare in the Age of Cyber Conflict*. Routledge.

Gorden, Raymond L. 1952. "Interaction between Attitude and the Definition of the Situation in the Expression of Opinion." *American Sociological Review* 17(1): 58.

Graphika. 2020. "Operation Naval Gazing: Facebook Takes Down Inauthentic Chinese Network." *Graphika Technical Report*. Accessed: https://public-assets.graphika.com/reports/graphika_report_naval_gazing.pdf

Graphika and the Stanford Internet Observatory. 2020. "More-Troll Kombat: French and Russian Influence Operations Go Head to Head Targeting Audiences in Africa." *Joint Technical Report by Graphika and the Stanford Internet Observatory*. Accessed: https://public-assets.graphika.com/reports/graphika_stanford_report_more_troll_kombat.pdf

Gray, Gordon and Elmer B. Staats. "Foreign Relations of the United States, 1950–1955, The Intelligence Community, 1950–1955 - Office of the Historian," accessed November 11, 2020, https://history.state.gov/historicaldocuments/frus1950-55Intel/d181

Grimaldi, James V. "FCC's Net Neutrality Proposal Marred by Millions of Fake Comments." *The Wall Street Journal.* May 6, 2021. Accessed: https://www.wsj.com/articles/fccs-net-neutrality-proposal-marred-by-millions-of-fake-comments-11620309351

Gross, Michael L., Daphna Canetti, and Dana R. Vashdi. 2016. "The Psychological Effects of Cyber Terrorism." *Bulletin of the Atomic Scientists* 72(5): 284–91.

Gross Stein, Janice. 1988. "Building Politics into Psychology: The Misperception of Threat." *Political Psychology* 9(2): 245-271.

Grossman, Guy, Devorah Manekin, and Yotam Margalit. 2018. "How Sanctions Affect Public Opinion in Target Countries: Experimental Evidence From Israel." *Comparative Political Studies* 51(14): 1823–1857.

Grossman, Shelby, Elena Cryst, Renée DiResta, K.H., Carly Miller, Rajeev Sharma, Chase Small, and Julia Thompson. 2021. "My Heart Loves the Army: An Investigation into a Jordanian Disinformation Campaign on Facebook, TikTok, and Twitter." *Stanford Internet Observatory Report*. Accessed: https://cyber.fsi.stanford.edu/io/news/jordanian-disinformation-campaign

Grossman, Shelby, Fazil Alp Akis, Ayça Alemdaroğlu, Josh A. Goldstein, and Katie Jonsson. 2020. "Political Retweet Rings and Compromised Accounts: A Twitter Influence

Operation Linked to the Youth Wing of Turkey's Ruling Party." *Stanford Internet Observatory Report*. Accessed: https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/20200611_turkey_report.pdf

Grossman, Shelby, Gil Baram, Josh A. Goldstein, and Carly Miller. 2021. "Staying Current; an Investigation into a Now-Suspended Facebook Network Promoting the Palestinian Democratic Reform Current." *Stanford Internet Observatory Report*. Accessed: https://cyber.fsi.stanford.edu/io/publication/staying-current-investigation-now-suspended-facebook-network-promoting-palestinian

Grossman, Shelby, Katie Jonsson, Nicholas Lyon, and Lydia Sizer. "Slanted Narratives, Social Media, and Foreign Influence in Libya." *Working Paper.*

Grossman, Shelby and Khadeja Ramali. 2020. "Outsourcing Disinformation." *Lawfare*. Accessed: https://www.lawfareblog.com/outsourcing-disinformation

Guisinger, Alexandra and Elizabeth N. Saunders. 2017. "Mapping the Boundaries of Elite Cues: How Elites Shape Mass Opinion Across International Issues." *International Studies Quarterly* 61(2): 425-441.

Hagey, Keach, Emily Glazer, and Rob Barry. "How Russia Today Skirts High-Tech Blockade to Reach U.S. Readers." *The Wall Street Journal.* October 7, 2020. Accessed: https://www.wsj.com/articles/how-russia-today-skirts-high-tech-blockade-to-reach-u-s-readers-11602078094

Hayes, Danny and Matt Guardino. 2011. "The Influence of Foreign Voices on U.S. Public Opinion," *American Journal of Political Science* 55(4): 831-851

Hedgecock, Kathryn, Lauren Sukin, and Leah Matchett. "The Distinctiveness of Public Response to Cyber and Kinetic Attacks." *Working Paper*.

Herman, Michael. 1996. *Intelligence Power in Peace and War.* Cambridge: Cambridge University Press.

Heuer, Richards J. 1981. "Strategic Deception and Counterdeception: A Cognitive Process Approach." *International Studies Quarterly* 25(2): 294-327.

Heuer, Richards Jr. 1990. "Assessing Soviet Influence Operations." *The Works of Richard Heuer, Jr.*

Holden-Reid, Brian. 2011. "The American Civil War And The Strategy Of Attrition, 1861–65." *The RUSI Journal*. 156(3): 88-95.

Holland, Max. 2006. "The Propagation and Power of Communist Security Services *Dezinformatsiya*." *International Journal of Intelligence and Counterintelligence* 19(1): 1-31.

Holsti, Ole R. 1992. “Public Opinion and Foreign Policy: Challenges to the Almond-Lippmann Consensus.” *International Studies Quarterly* 36(4): 439-466.

Holsti, Ole R. 2004. *Public Opinion and American Foreign Policy*. Rev. Analytical Perspectives on Politics. Ann Arbor: University of Michigan Press.

Horowitz, Michael and Stephen Rosen. 2005. “Evolution or Revolution?” *Journal of Strategic Studies* 28(3): 437-448.

Houck, Shannon C., James McFarland, Laura V. Machia, and Lucian Gideon Conway III. 2019. “When Beliefs Lead to (Im)Moral Action: How Believing in Torture’s Effectiveness Shapes the Endorsement of Its Use.” *Political Psychology* 40(6): 1315–1339.

Hovland, Carl I. and Walter Weiss. 1951. “The Influence of Source Credibility on Communication Effectiveness.” *Public Opinion Quarterly* 15: 635-650.

Howard, Philip N. 2020. *Lie Machines: How to Save Democracy from Troll Armies, Deceitful Robots, Junk News Operations, and Political Operatives*. New Haven: Yale University Press.

Howard, Philip N., Bharath Ganesh, Dimitra Liotsiou, John Kelly, Camille Francois. 2019. “The IRA, Social Media, and Political Polarization in the United States, 2012-2018.” Oxford Internet Institute Paper. Accessed: https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1004&context=senatedocs

Hundley, Lindsay, Renee DiResta, Josh A. Goldstein, Shelby Grossman, Cooper Reed, Adriana Stephan, and Julia Thomas. 2021. “One Face, Many Names: An Investigation into Fake NGOs and Media Outlets Linked to Harouna Douamba on and off Facebook.” *Stanford Internet Observatory Report*.

Hyde, Susan D. 2020. “Democracy’s backsliding in the international environment.” *Science* 369: 1192-1196.

Jack, Caroline. 2017. “Lexicon of Lies: Terms for Problematic Information.” *Data & Society Report*. Accessed: https://datasociety.net/pubs/oh/DataAndSociety_LexiconofLies.pdf

Jamieson, Kathleen Hall. 2020. *Cyberwar: How Russian Hackers and Trolls Helped Elect a President: What We Don’t, Can’t, and Do Know.* 2020. Oxford University Press.

Jentleson, Bruce W. 1992. “The Pretty Prudent Public: Post Post-Vietnam American Opinion on the Use of Military Force.” *International Studies Quarterly* 36(1): 49-73.

Jentleson, Bruce W. and Rebecca L. Britton. 1998. “Still Pretty Prudent: Post-Cold War American Public Opinion on the Use of Military Force.” *Journal of Conflict Resolution* 42(4): 395-417.

Jensen, Benjamin, and Brandon Valeriano. 2019. "What Do We Know about Cyber Escalation? Observations from Simulations and Surveys." *Atlantic Council: Scowcroft Center for Strategy and Security.*

Jervis, Robert. 1976. *Perception and Misperception in International Politics*. Princeton University Press.

Jervis, Robert. 1978. "Cooperation under the Security Dilemma." *World Politics* 30(2): 167-214.

Jervis, Robert. 2018. "Intelligence and International Politics," in *The Oxford Handbook of International Security* by Robert Jervis, Alexandra Gheciu and William C. Wohlforth, ed. Oxford University Press. pp. 515-530.

Jervis, Robert. 2017. *How Statesmen Think: The Psychology of International Politics*, *How Statesmen Think*. Princeton University Press.

Johnson, A. Ross. "Managing Media Influence Operations: Lessons from Radio Free Europe/Radio Liberty." *International Journal of Intelligence and Counterintelligence.* 31(4): 681-701.

Johnson, A. Ross. 2010. *Radio Free Europe and Radio Liberty: The CIA Years and Beyond.* Washington, DC: Wilson Center Press.

Johnson, Dominic D. P. 2020. *Strategic Instincts: The Adaptive Advantages of Cognitive Biases in International Politics*. Princeton Studies in International History and Politics. Princeton, New Jersey: Princeton University Press.

Johnson, Loch. 1989. "Covert Action and Accountability: Decision-Making for America's Secret Foreign Policy." *International Studies Quarterly* 33(1): 81-109.

Johnson, Loch. 2007. *Strategic Intelligence*, Volume 2: Intelligence and the Quest for Security. Westport, Conn.: Praeger Security International.

Joint Chiefs of Staff, DOD Dictionary of Military and Associated Terms, April 2019. Accessed: https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/dictionary.pdf

Joint Publication 3-13 "Information Operations" 2014. Accessed: https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_13.pdf

Jones, Bryan D. 1999. "Bounded Rationality." *Annual Review of Political Science* 2(1): 297-321.

Jones, Marc Owen. 2019. "Propaganda, Fake News, and Fake Trends: The Weaponization of Twitter Bots in the Gulf Crisis." *International Journal of Communication* 13: 1389-1415.

Joseph, Michael F. and Michael Poznansky. 2018. “Media Technology, Covert Action, and the Politics of Exposure.” *Journal of Peace Research* 55(3): 320–335.

Jost, Tyler and Joshua D. Kertzer. “Armies and Influence: Public Deference to Foreign Policy Elites.” *Working Paper.*

Jowett, Garth S. and Victoria O’Donnell. 2012. *Propaganda and Persuasion: Fifth Edition.* Sage Publications.

Kahn, Herman. 1965. *On Escalation: Metaphors and Scenarios*. London: Pall Mall Press.

Kalathil, Shanthi. 2020. “The Evolution of Authoritarian Digital Influence: Grappling with the New Normal.” *PRISM* 9(1): 33-50.

Kam, Cindy. 2005. “Who Toes the Party Line? Cues, Values, and Individual Differences.” *Political Behavior* 27(2): 163–82.

Kaminska, Monica. 2021. “Restraint under conditions of uncertainty: Why the United States tolerates cyberattacks.” *Journal of Cybersecurity* 7(1): 1-15.

Kay, Sean, and Patrick Cronin. “Bridging America’s Foreign Policy Elite-Main Street Divide.” *War on the Rocks*, August 11, 2016. Accessed: https://warontherocks.com/2016/08/bridging-americas-foreign-policy-elite-main-street-divide/

Kaysen, Carl. 1990. “Is War Obsolete?” *International Security* 14(4): 42-64.

Kello, Lucas. 2013. “The Meaning of the Cyber Revolution: Perils to Theory and Statecraft.” *International Security*. 38(2): 7-40.

Kelton, Mark. “Why ‘Unprecedented’ Is Never the Right Word for Russian Election-Meddling,” June 18, 2017, *The Cipher Brief*. Accessed: https://www.thecipherbrief.com/column_article/why-unprecedented-is-never-the-right-word-for-russian-election-meddling

Kennedy, Paul. 1987. *The Rise and Fall of the Great Powers.* Penguin.

Kent, Daniel, David Hirshleifer, and Avanidhar Subrahmanyam. 1998. “Investor Psychology and Security Market Under- and Overreactions.” *The Journal of Finance* 53(6): 1839-1885.

Kertzer, Joshua D. and Thomas Zeitzoff. 2017. “A Bottom-Up Theory of Public Opinion about Foreign Policy.” *American Journal of Political Science* 61(3): 543–58.

Kinetz, Erika. “Army of fake fans boosts China’s messaging on Twitter.” *The Associated Press*. May 28, 2021. Accessed: https://apnews.com/article/asia-pacific-china-europe-middle-east-government-and-politics-62b13895aa6665ae4d887dcc8d196dfc

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2017. "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument." *American Political Science Review* 111(3): 484-501.

Knopf, Jeffrey W. 1998. "How Rational is "The Rational Public"? Evidence from U.S. Public Opinion on Military Spending." *Journal of Conflict Resolution* 42(5): 544-571.

Krepinevich, Andrew F. 1994. "Cavalry to Computer: The Pattern of Military Revolutions." *The National Interest*. Accessed: https://nationalinterest.org/article/cavalry-to-computer-the-pattern-of-military-revolutions-848?page=0%2C2

Kreps, Sarah. 2010. "Elite Consensus as a Determinant of Alliance Cohesion: Why Public Opinion Hardly Matters for NATO-Led Operations in Afghanistan." *Foreign Policy Analysis* 6(3): 191–215.

Kreps, Sarah. 2020. *Social Media and International Relations*. Cambridge University Press.

Kreps, Sarah, and Debak Das. 2017. "Warring from the Virtual to the Real: Assessing the Public's Threshold for War over Cyber Security." *Research & Politics* 4(2).

Kreps, Sarah, and Jacquelyn Schneider. 2019. "Escalation Firebreaks in the Cyber, Conventional, and Nuclear Domains: Moving beyond Effects-Based Logics." *Journal of Cybersecurity* 5(1): 1-11.

Kreps, Sarah and Stephen Roblin. 2019. "Treatment format and external validity in international relations experiments." *International Interactions* 45(3): 576-594.

Kriner, Douglas, and Graham Wilson. 2016. "The Elasticity of Reality and British Support for the War in Afghanistan." *The British Journal of Politics and International Relations* 18(3): 559–80.

Kuklinski, James H., and Norman L. Hurley. 1994. "On Hearing and Interpreting Political Messages: A Cautionary Tale of Citizen Cue-Taking." *The Journal of Politics* 56(3): 729–51.

Kumkale, G. Tarcan, Dolores Albarracín, and Paul J. Seignourel. 2010. "The Effects of Source Credibility in the Presence or Absence of Prior Attitudes: Implications for the Design of Persuasive Communication Campaigns." *Journal of Applied Social Psychology* 40(6): 1325–1356.

Kux, Dennis. 1985. "Soviet Active Measures and Disinformation: Overview and Assessment." *Parameters* 15(4).

Larson, Eric V. and Bogdan Savych. 2015. *American Public Support for U.S. Military Operations from Mogadishu to Baghdad*. Santa Monica, CA: RAND Corporation.

Larson, Eric V. Richard E. Darilek, Daniel Gibran, Brian Nichiporuk, Amy Richardson, Lowell H. Schwartz, and Cathryn Quantic Thurston. 2009. *Foundations of Effective Influence Operations: A Framework for Enhancing Army Capabilities*, RAND Corporation Monograph Series.

Le Roux, Jean. 2021. "Paid influencers and sockpuppets drove social media campaign targeting South African banks." *Medium: DFRLab.* Accessed: https://medium.com/dfrlab/paid-influencers-and-sockpuppets-drove-social-media-campaign-targeting-south-african-banks-28bd04cd343

Lee, Dave. "The tactics of a Russian troll farm." *BBC.* February 16, 2018. Accessed: https://www.bbc.com/news/technology-43093390

Lenz, Gabriel S. 2009. "Learning and Opinion Change, Not Priming: Reconsidering the Priming Hypothesis." *American Journal of Political Science* 53(4): 821–37.

Levendusky, Matthew S. and Michael C. Horowitz. 2012. "When Backing Down Is the Right Decision: Partisanship, New Information, and Audience Costs." *The Journal of Politics* 74(2): 323-338.

Levy, Jack S. and William R. Thompson. 2011. *The Arc of War: Origins, Escalation, and Transformation*. Chicago, Ill.: University of Chicago Press.

Libicki, Martin C. 2017. "The Convergence of Information Warfare." *Strategic Studies Quarterly* 11(1): 49-65.

Libicki, Martin C. 2018. "Expectations of Cyber Deterrence." *Strategic Studies Quarterly* 12(4): 44-57.

Lin, Herbert. 2016. "Attribution of Malicious Cyber Incidents: From Soup to Nuts." *Journal of International Affairs* 70(1): 75-137.

Lin, Herbert. 2019. "The existential threat from cyber-enabled information warfare." *Bulletin of the Atomic Scientists* 75(4): 187-196.

Lin, Herbert and Jaclyn Kerr. 2019. "On Cyber-Enabled Information Warfare and Information Operations." In Oxford Handbook of Cybersecurity.

Lin-Greenberg, Erik and Theo Milonopoulos. 2021. "Private Eyes in the Sky: Emerging Technology and the Political Consequences of Eroding Government Secrecy." *Journal of Conflict Resolution* 65(6): 1067-1097.

Lindsay, Jon R. 2015. "Tipping the Scales: The Attribution Problem and the Feasibility of Deterrence against Cyberattack." *Journal of Cybersecurity* 1(1): 53-67.

Linvill, Darren L., Brandon C. Boatwright, Will J. Grant, and Patrick L. Warren. 2019. "'THE RUSSIANS ARE HACKING MY BRAIN!' investing Russia's internet research agency twitter tactics during the 2016 United States presidential campaign." *Computers in Human Behavior* 99: 292-300.

Linvill, Darren L. and Patrick L. Warren. 2020. "Engaging with Others: How the IRA Coordinated Information Operation Made Friends." *The Harvard Kennedy School Misinformation Review* 1(2): 1-14.

Linvill, Darren L. and Patrick L. Warren. 2020. "Troll Factories: Manufacturing Specialized Disinformation on Twitter." *Political Communication*: 1-21.

Linvill, Darren and Patrick Warren. "Foreign Influence Operations Don't Need to Succeed to Be Effective," October 23, 2020. *Lawfare*. Accessed: https://www.lawfareblog.com/foreign-influence-operations-dont-need-succeed-be-effective.

Lippmann, Walter. 1955. *Essays in the Public Philosophy*. New York: New American Library.

Lombardo, Johannes R. 1999. "A Mission of Espionage, Intelligence and Psychological Operations: The American Consulate in Hong Kong, 1949–64." *Intelligence and National Security* 14(4): 64-81.

Lowenthal, Mark M. 2012. *Intelligence: From Secrets to Policy.* CQ Press.

Lukito, Josephine. 2019. "Coordinating a Multi-Platform Disinformation Campaign: Internet Research Agency Activity on Three U.S. Social Media Platforms, 2015 to 2017." *Political Communication* 37(2): 238-255.

Lukito, Josephine. 2020. "Coordinating a Multi-Platform Disinformation Campaign: Internet Research Agency Activity on Three U.S. Social Media Platforms, 2015 to 2017." *Political Communication* 37(2): 238-255.

Lukito, Josephine, Jiyoun Suk, Yini Zhang, Larissa Doroshenko, Sang Jung Kim, Min-Hsin Su, Yiping Xia, Deen Freelon, and Chris Wells. 2019. "The Wolves in Sheep's Clothing: How Russia's Internet Research Agency Tweets Appeared in U.S. News as Vox Populi." *The International Journal of Press/Politics* 25(2): 196-216.

Lupia, Arthur and Mathew D. McCubbins. 1998. *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* Cambridge University Press.

Lynch, Marc. "Media, Old and New." *The Arab Uprisings Explained*. Edited by Marc Lynch, Columbia University Press. pp. 93-109.

M.A., Renée DiResta, Josh A. Goldstein, and Shelby Grossman. "Middle East Influence Operations: Observations Across Social Media Takedowns." Paper prepared for the

Project on Middle East Political Science's 2021 Conference, "Digital Activism and Authoritarian Adaptation in the Middle East."

Madrigal, Alexis C. "Russia's Troll Operation Was Not That Sophisticated." *The Atlantic*. February 19, 2018. Accessed: https://www.theatlantic.com/technology/archive/2018/02/the-russian-conspiracy-to-commit-audience-development/553685/

Maoz, Ifat, Andrew Ward, Michael Katz, and Lee Ross. 2002. "Reactive devaluation of an "Israeli" vs."Palestinian" peace proposal." *Journal of Conflict Resolution* 46(4): 515-546.

Marsh, Catherine. 1985. "Back on the Bandwagon: The Effect of Opinion Polls on Public Opinion." *British Journal of Political Science* 15(1): 51–74.

Martin, Diego A., Jacob N. Shapiro, and Julia Ilhardt. "Trends in Online Foreign Influence Efforts: Version 2.0." *Working Paper*.

Matthews, John P. C. 2003. "The West's Secret Marshall Plan for the Mind." *International Journal of Intelligence and Counterintelligence* 16(3): 409-427.

Matthews, Miriam, Christopher Paul, David Schulker, and David Stebbins. 2020. "Frameworks for Assessing USEUCOM Efforts to Inform, Influence, and Persuade." *RAND Report*.

Mazarr, Michael J. 2018. "Understanding Deterrence." RAND Perspective. Accessed: https://www.jstor.org/stable/resrep19896.

Mazzetti, Mark, Adam Goldman, and Michael S. Schmidt. "Behind the Sudden Death of a $1 Billion Secret C.I.A. War in Syria," *The New York Times*, August 2, 2017. Accessed: https://www.nytimes.com/2017/08/02/world/middleeast/cia-syria-rebel-arm-train-trump.html

McDermott, Rose. 2004. *Political Psychology in International Relations*. Analytical Perspectives on Politics. Ann Arbor: The University of Michigan Press.

McGee, Luke. "The brazen arrest of a Belarusian activist has terrified dissidents all over the world." CNN. May 30, 2021. Accessed: https://www.cnn.com/2021/05/29/europe/belarus-arrest-ryanair-dissidents-unsafe-intl-cmd/index.html

McGuire, William J. 1969. "The Nature of Attitudes and Attitude Change." In *The Handbook of Social Psychology*, edited by Gardner Lindzey and Elliot Aronson, 2nd ed. Reading, MA: Addison-Wesley.

Mearsheimer, John J. 1994. "The False Promise of International Institutions." *International Security* 19(3): 5-49.

Mearsheimer, John J. 2011. *Why Leaders Lie: The Truth about Lying in International Politics*. New York, N.Y.: Oxford University Press.

Mermin, Jonathan. 1999. *Debating War and Peace: Media Coverage of US Intervention in the Post-Vietnam Era*. Princeton University Press.

Messing, Solomon and Sean J. Westwood. 2014. "Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online." *Communication Research* 41(8): 1042-1063.

Miller, Carly, Vanessa Molter, Isabella Garcia-Camargo, and Renée DiResta. 2020. "Sockpuppets Spin COVID Yarns: An Analysis of PRC-Attributed June 2020 Twitter takedown." *Stanford Internet Observatory Report.*

Miller, Maggie. "Majority of Americans concerned about potential foreign interference: poll." *The Hill*. October 2, 2020. Accessed: https://thehill.com/policy/cybersecurity/519411-majority-of-americans-concerned-about-potential-foreign-election

Milner, Helen V. and Dustin Tingley. 2015. *Sailing the Water's Edge.* Princeton University Press.

Mondak, Jeffery J. 1993. "Source Cues and Policy Approval: The Cognitive Dynamics of Public Support for the Reagan Agenda." *American Journal of Political Science* 37(1): 186-212.

Morgenthau, Hans J. 1978. *Politics Among Nations: The Struggle for Power and Peace.* New York: Alfred A. Knopf.

Mueller, John E. 1971. "Trends in Popular Support for the Wars in Korea and Vietnam." *American Political Science Review* 65(2): 358-375.

Mueller, John E. 1973. *War, Presidents, and Public Opinion*. New York: John Wiley & Sons.

Munck, Gerardo L. and Richard Snyder. 2007. "Debating the Direction of Comparative Politics: An Analysis of Leading Journals." *Comparative Political Studies* 40(1): 5-31.

Murray, Shoon. 2014. "Broadening the Debate about War: The Inclusion of Foreign Critics in Media Coverage and Its Potential Impact on US Public Opinion." *Foreign Policy Analysis* 10(4): 329-350.

Myrick, Rachel. 2020. "Why So Secretive? Unpacking Public Attitudes toward Secrecy and Success in US Foreign Policy." *Journal of Politics* 82(3): 828-843.

Nakashima, Ellen, Amy Gardner, and Aaron C. Davis. "FBI links Iran to online hit list targeting top officials who've refuted Trump's election fraud claims." *The Washington Post.* December 22, 2020. Accessed: https://www.washingtonpost.com/national-security/iran-election-fraud-violence/2020/12/22/4a28e9ba-44a8-11eb-a277-49a6d1f9dff1_story.html

Nakashima, Ellen. "U.S. Cyber Command operation disrupted Internet access of Russian troll factory on day of 2018 midterms." *The Washington* Post. February 27, 2019. Accessed: https://www.washingtonpost.com/world/national-security/us-cyber-command-operation-disrupted-internet-access-of-russian-troll-factory-on-day-of-2018-midterms/2019/02/26/1827fc9e-36d6-11e9-af5b-b51b7ff322e9_story.html

National Geographic, Gallup, and the Council on Foreign Relations. 2019. "U.S. Adults' Knowledge About the World." *Joint Report.* p. 4. Accessed: https://cdn.cfr.org/sites/default/files/report_pdf/NatGeo_CFR_US%20Knoweldge.pdf

Neuman, W. Russell. 1986. *The Paradox of Mass Politics: Knowledge and Opinion in the American Electorate*. Cambridge, Mass: Harvard University Press.

Nimmo, Ben. 2020. "The Breakout Scale: Measuring the Impact of Influence Operations," *Brookings Report*. Accessed: https://www.brookings.edu/research/the-breakout-scale-measuring-the-impact-of-influence-operations/

Nincic, Miroslav. "A Sensible Public: New Perspectives on Popular Opinion and Foreign Policy." *Journal of Conflict Resolution* 36(4): 772-789.

Nye, Joseph S. 2004. "Soft Power and American Foreign Policy." *Political Science Quarterly* 119(2): 255-270.

Nye, Joseph S. 2019. "Protecting Democracy in an Era of Cyber Information War." *Belfer Center Paper*.

O'Driscoll, Cian. 2015. "Rewriting the Just War Traditions: Just War in Classical Greek Political Thought and Practice." *International Studies Quarterly* 29: 1-10.

Oneal, John R., Brian Lian, and James H. Joyner Jr. 1996. "Are the American People "Pretty Prudent"? Public Responses to U.S. Uses of Force, 1950-1988." *International Studies Quarterly* 40(2): 261-280.

Ördén, Hedvig and James Pamment. 2021. "What Is So Foreign About Foreign Influence Operations?" *Carnegie Endowment for International Peace.* Accessed: https://carnegieendowment.org/2021/01/26/what-is-so-foreign-about-foreign-influence-operations-pub-83706

Orr, Lilla V. and Gregory A. Huber. 2020. "The Policy Basis of Measured Partisan Animosity in the United States." *American Journal of Political Science* 64(3): 569-586.

Ossoff, William. 2021. "Hacking the Domaine Réservé: The Rule of Non-Intervention and Political Interference in Cyberspace." *Harvard International Law Journal* 62(1): 295-323.

Pacepa, Mihai and Ronald J. Rychlak, 2013. *Disinformation: Former Spy Chief Reveals Secret Strategies for Undermining Freedom, Attacking Religion, and Promoting Terrorism.* WND Books.

Page, Benjamin I., and Lawrence R. Jacobs. 2006. "A Disconnect Between Policy Makers and the Public?" In *The Foreign Policy Disconnect: What Americans Want from Our Leaders but Don't Get*, by Benjamin I. Page and Marshall M. Bouton. Chicago, Ill.: University of Chicago Press.

Page, Benjamin I. and Robert Y. Shapiro. 1983. "Effects of Public Opinion on Policy." *The American Political Science Review* 77(1): 175-190.

Page, Benjamin I. and Robert Y. Shapiro. 1992. *The Rational Public: Fifty Years of Trends in Americans' Policy Preferences*. American Politics and Political Economy. Chicago: University of Chicago Press.

Pamment, James and Alicia Wanless. 2019. "How Do You Define a Problem Like Influence?" *Journal of Information Warfare* 18(3): 1-14.

Pamment, James, Howard Nothhaft, and Alicia Fjällhed. 2019. "Countering Information Influence Activities: A Handbook for Communicators." *Swedish Civil Contingencies Agency Report*.

Pattison, James. 2015."The Ethics of Arming Rebels." *Ethics & International Affairs* 29(4): 455–471.

Paul, T.V., James J. Wirtz, and Michel Fortmann. 2004. *Balance of Power: Theory and Practice in the 21st Century*. Stanford, California: Stanford University Press.

Pearson, Lionel. 1936. "Propaganda in the Archidamian War." *Classical Philology* 31(1) 33–52.

Perlroth, Nicole. "Russians Who Pose Election Threat Have Hacked Nuclear Plants and Power Grid." *The New York Times*. October 23, 2020. Accessed: https://www.nytimes.com/2020/10/23/us/politics/energetic-bear-russian-hackers.html

Pew Research Center. "Americans more likely to get news on digital devices from new websites, apps and search engines than from social media." *Pew Research Center.* January 11, 2021. Accessed: pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices/ft_2021-01-12_socialmedia_03/

Pijnenburg Muller, Lilly and Tim Stevens. 2017. "Upholding the NATO Cyber Pledge." *Norwegian Institute of International Affairs Policy Brief.*

Pomirchy, Michael and Bryan Schonfeld. "Does Responsiveness Stop at the Water's Edge? Examining the Foreign Policy Electoral Connection" pp. 16-18. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3436890

Poole, Steven. "What's the difference between a troll and a sockpuppet?" *The Guardian.* February 23, 2018. Accessed: https://www.theguardian.com/books/2018/feb/23/troll-steven-poole-word-of-week

Pornpitakpan, Chanthika. 2004. "The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence." *Journal of Applied Social Psychology* 34(2): 243-281.

Potter, Philip B. K. and Matthew A. Baum. 2014. "Looking for Audience Costs in all the Wrong Places: Electoral Institutions, Media Access, and Democratic Constraint." *The Journal of Politics* 76(1): 167-181.

Poznansky, Michael. 2019. "Feigning Compliance: Covert Action and International Law." *International Studies Quarterly.* 63(1): 72-84.

Poznansky, Michael. 2020. "Revisiting plausible deniability." *Journal of Strategic Studies*: 1-23.

Poznansky, Michael and Evan Perkoski. 2018. "Rethinking Secrecy in Cyberspace: The Politics of Voluntary Attribution." *Journal of Global Security Studies*. 3(4): 402-416.

Pullin, Eric D. 2011. "'Money Does Not Make Any Difference to the Opinions That We Hold': India, the CIA, and the Congress for Cultural Freedom, 1951–58." *Intelligence and National Security* 26(2-3): 377-398.

Putnam, Robert D. 1988. "Diplomacy and Domestic Politics: The Logic of Two-Level Games." *International Organization* 42(3): 427-460.

Qualter, Terence H. 1962. *Propaganda and Psychological Warfare*. Studies in Political Science. New York: Random House.

Radsan, A. John. 2009. "An Overt Turn on Covert Action." *Saint Louis University Law Journal* 53(2): 485-552.

Rawnsley, Gary D. 1996. "Overt and Covert: The Voice of Britain and Black Radio Broadcasting in the Suez Crisis, 1956." *Intelligence and National Security* 11(3): 497-522.

Rid, Thomas. 2020. *Active Measures: The Secret History of Disinformation and Political Warfare*. Farrar, Straus and Giroux.

Rid, Thomas and Ben Buchanan. 2015. "Attributing Cyber Attacks." *Journal of Strategic Studies* 38(1-2): 4-37.

Robinson, Joshua. 2021. "Can Elites Escape Blame by Explaining Themselves? Suspicion and the Limited of Elite Explanations." *British Journal of Political Science*: 1-20.

Robinson, Linda, Todd C. Helmus, Raphael S. Cohen, Alireza Nader, Andrew Radin, Madeline Magnuson, and Katya Migacheva. 2019. "The Growing Need to Focus on Modern Political Warfare." *RAND Report*.

Rothschild, David, and Neil Malhotra. 2014. "Are Public Opinion Polls Self-Fulfilling Prophecies?" *Research & Politics* 1(2): 1-10.

Sagan, Paul and Tom Leighton. 2010. "The Internet & the future of news." *Daedalus* 139(2): 119-125.

Sagan, Scott D. and Benjamin A. Valentino. 2017. "Revisiting Hiroshima in Iran: What Americans Really Think about Using Nuclear Weapons and Killing Noncombatants." *International Security* 42(1): 41-79.

Sagar, Rahul. 2013. *Secrets and Leaks: The Dilemma of State Secrecy*. Princeton, NJ: Princeton University Press

Sánchez-Cuenca, Ignacio and Luis de la Calle. 2009. "Domestic Terrorism: The Hidden Side of Political Violence." *Annual Review of Political Science* (12)1: 31-49.

Sander, Bernie. 2019. "Democracy Under The Influence: Paradigms of State Responsibility for Cyber Influence Operations on Elections." *Chinese Journal of International Law.* 18(1): 1-56.

Sartori, Giovanni. 1970. "Concept Misformation in Comparative Politics." *The American Political Science Review* 64(4): 1033-1053.

Schatz, Edward and Renan Levine. 2010. "Framing, Public Diplomacy, and Anti-Americanism in Central Asia." *International Studies Quarterly* 54(3): 855-869.

Schleifer, Ron. 2006. "Psychological operations: A new variation on an age old art: Hezbollah versus Israel." *Studies in Conflict & Terrorism* 29(1): 1-19.

Schmidt, Brian C. "On the history and historiography of international relations." In: Carlsnaes W., Risse T. and Simmons BA (eds) *Handbook of International Relations.* Sage. pp. 3-28.

Schmitt, Michael N. 2017. "Grey Zones in the International Law of Cyberspace." *The Yale Journal of International Law*. 42(2): 1-21.

Schuessler, John M. 2015. *Deceit on the Road to War: Presidents, Politics and American Democracy*. Cornell Studies in Security Affairs. Ithaca: Cornell University Press.

Shandler, Ryan, Michael L. Gross, Sophia Backhaus, and Daphna Canetti. 2021. "Cyber Terrorism and Public Support for Retaliation: A Multi-Country Survey Experiment." *British Journal of Political Science*: 1-19.

Shandler, Ryan, Michael L. Gross, and Daphna Canetti. 2021. "A Fragile Public Preference for Cyber Strikes: Evidence from Survey Experiments in the United States, United Kingdom, and Israel." *Contemporary Security Policy* 42(2): 135-162.

Shapiro, Robert Y. and Benjamin I. Page. 1988. "Foreign Policy and the Rational Public." *Journal of Conflict Resolution* 32(2): 211–47.

Shearer, Elisa and Amy Mitchell. "News Use Across Social Media Platforms in 2020." *Pew Research Center.* January 12, 2021. Accessed: https://www.journalism.org/2021/01/12/news-use-across-social-media-platforms-in-2020/?utm_source=AdaptiveMailer&utm_medium=email&utm_campaign=21-01-12%20Social%20Media%20Update%20General%20Distro&org=982&lvl=100&ite=7690&lea=1696518&ctr=0&par=1&trk=

Shearlaw, Maeve. "Turkish journalists face abuse and threats online as trolls step up attacks," *The Guardian*, November 1, 2016. Accessed: https://www.theguardian.com/world/2016/nov/01/turkish-journalists-face-abuse-threats-online-trolls-attacks

Shimer, David. 2020. *Rigged: America, Russia and 100 Years of Covert Electoral Interference* HarperCollins Publishers.

Shires, James. 2020. "The Simulation of Scandal: Hack-and-Leak Operations, the Gulf States, and U.S. Politics," *Texas National Security Review* 3(4): 10-29.

Shulman, Stephen and Stephen Bloom. 2012. "The legitimacy of foreign intervention in elections: the Ukrainian response." *Review of International Studies.* 38(2): 445-471.

Silove, Nina. "Beyond the Buzzword: The Three Meanings of 'Grand Strategy.'" *Security Studies* 27(1): 27-57.

Silverman, Craig. "Why Do A Bunch Of Nigerian Twitter Influencers Want This Alleged Money Launderer To Go Free? They're Being Paid." *BuzzFeed News.* April 9, 2021. Accessed: https://www.buzzfeednews.com/article/craigsilverman/nigerian-twitter-campaign-alex-saab

Silverman, Craig, Jane Lytvynenko, and William Kung. "Disinformation For Hire: How a New Breed of PR Firms Is Selling Lies Online." January 6, 2020. *BuzzFeed News*. Accessed: https://www.buzzfeednews.com/article/craigsilverman/disinformation-for-hire-black-pr-firms

Smith, Alexander, Kristen Welker, Andrea Mitchell, and Abigail Williams. "Biden calls for de-escalation with Russia, following sanctions, proposes meeting with Putin." *NBC News.* April 15, 2021. Accessed: https://www.nbcnews.com/news/world/u-s-sanction-russia-alleged-election-interference-solarwinds-hack-n1264142

Smith, Victoria. 2020. “Mapping Worldwide Initiatives to Counter Influence Operations.” *Carnegie Endowment for International Peace*. Accessed: https://carnegieendowment.org/2020/12/14/mapping-worldwide-initiatives-to-counter-influence-operations-pub-83435

Sniderman, Paul M., Richard A. Brody, and Philip E. Tetlock. 1991. *Reasoning and Choice: Explorations in Political Psychology*. Cambridge: Cambridge University Press.

Starbird, Kate. “The Surprising Nuance Behind the Russian Troll Strategy.” *Medium*. October 20, 2018. Accessed: https://medium.com/s/story/the-trolls-within-how-russian-information-operations-infiltrated-online-communities-691fb969b9e4

Starbird, Kate, Ahmer Arif, and Tom Wilson. 2019. “Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations.” *Proceedings of the ACM on Human-Computer Interaction* 3: 1-26.

Stecula, Dominik A. and Mark Pickup. 2021. “How populism and conservative media fuel conspiracy beliefs about COVID-19 and what it means for COVID-19 behaviors.” *Research & Politics*: 1-9.

Steward, Leo G., Ahmer Arif, and Kate Starbird. 2018. “Examining trolls and polarization with a retweet network.” *MIS2, 2018.*

Stubbs, Jack. “Exclusive-Russian operation masqueraded as right-wing news site to target U.S. voters.” *Reuters*, October 1, 2020. Accessed: https://www.reuters.com/article/usa-election-russia-disinformation/exclusive-russian-operation-masqueraded-as-right-wing-news-site-to-target-u-s-voters-sources-idINL1N2GS0FY

Sulick, Michael J. 2012. *Spying in America: Espionage from the Revolutionary War to the Dawn of the Cold War*. Georgetown University Press.

Surowiecki, James. 2004. *The Wisdom of Crowds: Why the Many Are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. London: Little Brown.

Sweijs, Tim and Samo Zilincik. 2019. “Deterrence by Democracies in a Hybrid Context.” *Cross Domain Deterrence and Hybrid Conflict* Hague Centre for Strategic Studies. Accessed: https://www.jstor.org/stable/resrep24191.6

Sykes, Charles J. “Why Nobody Cares the President is Lying.” *The New York Times* February 4, 2017. Accessed: https://www.nytimes.com/2017/02/04/opinion/sunday/why-nobody-cares-the-president-is-lying.html

Tajfel, Henry, and John Turner. 1979. "Chapter 3 An Integrative Theory of Intergroup Conflict." In *The Social Psychology of Intergroup Relations*, edited by William G. Austin and Stephen Worchel.

Tetlock, Philip E. 2005. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, N.J.: Princeton University Press.

The National Archives (TNA), "Letter to the Embassy in regard to public opinion polls in America," March 19th, 1940, FO 371/24228/265-268.

Tomz, Michael. 2007. "Domestic Audience Costs in International Relations: An Experimental Approach." *International Organization* 61(4): 821-840.

Tomz, Michael and Jessica L. P. Weeks. 2020. "Public Opinion and Foreign Electoral Intervention." *American Political Science Review* 114(3): 856-873.

Tomz, Michael, Jessica L.P. Weeks, and Keren Yarhi-Milo. 2020. "Public Opinion and Decisions About Military Force in Democracies." *International Organization* 74(1): 119-143.

Tomz, Michael and Jessica L. P. Weeks. "Rhetoric, Partisanship, and Public Reactions to Foreign Electoral Intervention." *Working Paper.* (n.d.-a)

Tomz, Michael and Jessica L. P. Weeks. "The Domestic Political Costs of Soliciting Foreign Electoral Intervention." *Working Paper.* (n.d.-b)

Turcotte, Jason, Chance York, Jacob Irving, Rosanne M. Scholl, and Raymond J. Pingree. 2015. "News Recommendations from Social Media Opinion Leaders: Effects on Media Trust and Information Seeking." *Journal of Computer-Mediated Communication* 20: 520-535.

Torngren, Gustaf, and Henry Montgomery. 2004. "Worse Than Chance? Performance and Confidence Among Professionals and Laypeople in the Stock Market." *Journal of Behavioral Finance* 5(3): 148–53.

Twitter Safety. "New disclosure to our archive of state-backed information operations." *Twitter Blog*. December 20, 2019. Accessed: https://blog.twitter.com/en_us/topics/company/2019/new-disclosures-to-our-archive-of-state-backed-information-operations.html

United States Department of State. 1981. "Soviet Active Measures: Forgery, Disinformation, Political Operations." Special Report No. 88. Accessed: https://www.hsdl.org/?abstract&did=807615

United States Department of State. 1986. *Active Measures: Process of Anti-U.S. Disinformation and Propaganda Campaigns*. Department of State Publication 9630. Accessed: http://insidethecoldwar.org/sites/default/files/documents/Soviet%20Active%20Measures

%20Substance%20and%20Process%20of%20Anti-US%20Disinformation%20August%201986.pdf

United States Department of State. 2020. "Pillars of Russia's Disinformation and Propaganda System." *Global Engagement Center Special Report.* Accessed: https://www.state.gov/wp-content/uploads/2020/08/Pillars-of-Russia%E2%80%99s-Disinformation-and-Propaganda-Ecosystem_08-04-20.pdf

Usdin, Steve. "When a Foreign Government Interfered in a U.S. Election—to Reelect FDR." *Politico Magazine*. January 16, 2017. Accessed: https://www.politico.com/magazine/story/2017/01/when-a-foreign-government-interfered-in-a-us-electionto-reelect-fdr-214634

Van Evera, Stephen. 1990. "Primed for Peace: Europe after the Cold War." *International Security* 15(3): 7-57.

Vavra, Shannon. "Election interference efforts have shifted, NSA and Cyber Command election threats lead say." *CyberScoop.* August 7, 2020. Accessed: https://www.cyberscoop.com/election-interference-2020-presidential-elections-nsa-cyber-comand-russia-china/

Verba, Sidney, Richard A. Brody, Edwin B. Parker, Norman H. Nie, Nelson W. Polsby, Paul Ekman, and Gordon S. Black. 1967. "Public Opinion and the Vietnam War" *The American Political Science Review* 61(2): 317-333

Volume I of the Report On the Investigation Into Russian Interference In The 2016 Presidential Election, Special Counsel Robert S. Mueller III. Accessed: https://www.justice.gov/archives/sco/file/1373816/download

Walker, Christopher. 2018. "What Is 'Sharp Power'?" *Journal of Democracy* 29(3): 9-23.

Walker, Christopher and Jessica Ludwig. 2017. "Sharp Power: Rising Authoritarian Influence" *National Endowment for Democracy Report.*

Wallenius, Claes and Sofia Nilsson. 2019. "A Lack of Effect Studies and of Effects: The Use of Strategic Communication in the Military Domain." *International Journal of Strategic Communication*. 13(5): 404-417.

Waller, J. Michael. 2009. "The American Way of Propaganda: Learning from the Founding Fathers." *Strategic Influence: Public Diplomacy, Counterpropaganda, and Political Warfare* Washington, D.C.: The Institute of World Politics Press.

Wallis, Jake, Tom Uren, Elise Thomas, Albert Zhang, Samantha Hoffman, Lin Li, Alex Pascoe, and Danielle Cave. 2020. "Retweeting through the great firewall." *Australian Strategic Policy Institute Report No. 33/2020.*

Walt, Stephen M. 1987. *The Origins of Alliances*. Cornell Studies in Security Affairs. Ithaca, New York: Cornell University Press.

Walton, Calder. 2018. "Historical Amnesia: British and U.S. Intelligence, Past and Present," *Secrecy and Society* 2(1): 1-30.

Waltz, Kenneth N. 1959. *Man, the State, and War: A Theoretical Analysis*. Topical Studies in International Relations; New York.

Ward, Clarissa, Katie Polglase, Sebastian Shukla, Gianluca Mezzofiore, and Tim Lister. "Russian election meddling is back – via Ghana and Nigeria – in your feeds." *CNN*. April 11, 2020. Accessed: https://www.cnn.com/2020/03/12/world/russia-ghana-troll-farms-2020-ward/index.html

Wardle, Claire. 2020. "The Media Has Overcorrected on Foreign Influence." *Lawfare.* Accessed: https://www.lawfareblog.com/media-has-overcorrected-foreign-influence

Wardle, Claire and Hossein Derakhsahn. 2018. "Thinking about 'information disorder': formats of misinformation, disinformation and mal-information." In Ireton, C., & Posetti, J. (Eds.) Journalism, 'fake news', & disinformation: handbook for journalism education and training. UNESCO. pp. 43-53.

Warner, Michael. 2009. "Intelligence Theory: Key Questions and Debates," in Mark Phythian, Peter Gill, and Stephen Marrin, ed., *Intelligence Theory: Key Questions and Debates*. Studies in Intelligence Series. London: Routledge. pp. 16-17.

Warren, Mark E. 2014. "Accountability and Democracy." *The Oxford Handbook of Public Accountability*. pp. 39-54.

Weedon, Jen, William Nuland, and Alex Stamos. 2017. "Information Operations and Facebook." *Facebook*. Accessed: https://i2.res.24o.it/pdf2010/Editrice/ILSOLE24ORE/ILSOLE24ORE/Online/_Oggetti_Embedded/Documenti/2017/04/28/facebook-and-information-operations-v1.pdf

Weeks, Jessica L. P. 2008. "Autocratic Audience Costs: Regime Type and Signaling Resolve." *International Organization* 62(1): 35-64.

Wigell, Mikael. 2019. "Hybrid Interference as a Wedge Strategy: A Theory of External Interference in Liberal Democracy." *International Affairs* 95(2): 255–275.

Wildavsky, Aaron. 1966. "The Two Presidencies." *Society* 35(2): 23-31.

Wilkin, Bernard and Maude Williams. 2017. "German Wartime Anglophobic Propaganda in France, 1914–1945." *War in History* 24(1): 28–43.

Williams, Daniel. 2020. “Motivated ignorance, rationality, and democratic politics.” *Synthese*: 1-21.

Williamson, Scott. “Proxies and Plausible Deniability” *Working Paper.*

Wohlforth, William C., Richard Little, Stuart J. Kaufman, David Kang, Charles A. Jones, Victoria Tin-Bor Hui, Arthur Eckstein, Daniel Deudney, and William L. Brenner. 2007. “Testing Balance-of-Power Theory in World History.” *European Journal of International Relations* 13(2): 155-185.

Woolley, Samuel. 2020. “Bots and Computational Propaganda: Automation for Communication and Control,” in Nathaniel Persily and Joshua Tucker, ed., *Social Media and Democracy: The State of the Field, Prospects for Reform*. Cambridge: Cambridge University Press. pp. 89–110.

Yadav, Kamya. 2020. “Countering Influence Operations: A Review of Policy Proposals Since 2016.” *Carnegie Endowment for International Peace Report*.

Yaffa, Joshua. “Is Russian Meddling as Dangerous As We Think?” *The New Yorker*. September 7, 2020. Accessed: https://www.newyorker.com/magazine/2020/09/14/is-russian-meddling-as-dangerous-as-we-think

Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.

Zaller, John. 2012. “What Nature and Origin Leaves Out.” *Critical Review* 24(4): 569–642.

Zegart, Amy B. 1999. *Flawed by Design: The Evolution of the CIA, JCS, and NSC*. Stanford University Press.

Zeitzoff, Thomas. 2018. “Does Social Media Influence Conflict? Evidence from the 2012 Gaza Conflict.” *Journal of Conflict Resolution* 62(1): 29-63.

Zhang, Yini, Josephine Lukito, Min-Hsin Su, Jiyoun Suk, Yiping Xia, Sang Jun Kim, Larissa Doroshenko, and Chris Wells. 2021. “Assembling the Networks and Audiences of Disinformation: How Successful Russian IRA Twitter Accounts Built Their Followings, 2015-2017.” *Journal of Communication* 71(2): 305-331.