THE UNIVERSITY OF CHICAGO

ESSAYS ON COMPUTATIONAL TECHNOLOGIES IN INTERNATIONAL RELATIONS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF POLITICAL SCIENCE

BY
PETER BRENNAN CUPPERNULL

CHICAGO, ILLINOIS
JUNE 2024

Contents

List of Tables

List of Figures

Acknowledgements

I am deeply grateful to have found my intellectual home in Hyde Park, and I could not have written this dissertation without the support of numerous individuals in that community. First, I am thankful to my committee members for their guidance and support. Rochelle Terman's commitment to my work far exceeded that of an advisor's job description. Rochelle provided me with my first co-authorship opportunity, helped me navigate the frustration and anger of failing a comprehensive exam, introduced me to the people who helped me land my internship and job, and even funded much of this research. More important than the impact she had on this dissertation, however, is the impact she had on my life. Rochelle "let slip" to the department that I moonlight as a musician, and her unreserved acceptance of that side of my life has encouraged me to be a more authentic version of myself.

Austin Carson's unending willingness to engage with my ideas and push me when I needed it has made this work better; many of the connections I make in this dissertation to the broader field are due to his guidance. Like Rochelle, he has supported me as more than just a student. Finally, James Evans inspired me to take the road less traveled in my research. He encouraged me to pursue ideas that I care about and reminded me that there is an art behind the science.

Beyond my committee, I am grateful to Bobby Gulotty, Paul Poast, Anton Strezhnev, and John Mearsheimer for engaging with and contributing to my research. I would not have attended graduate school if not for the mentorship of Rich Rubenstein at George Mason University. Rich was the first person to take me seriously as a scholar, and his confidence in me made me more confident in myself.

My friendships with other students in the department were essential to the completion of this dissertation. Nobody spent more time helping me develop my ideas than Noah Schouela. He is a fierce friend and bakes delicious bread. Thank you to Adam Saxton, Elsy Gonzalez,

Evgenia Olimpieva, Ipek Cinar, Jenna Gibson, Josh Byun, Maddie Stevens, and Steven Boyd for your friendship and support. It has been a privilege to learn and grow with each of you, and it was only with your friendship that simultaneously navigating both graduate school and a pandemic was possible. This work also greatly benefited from the feedback of attendees of the Workshop on International Politics at the University of Chicago, the ISCS Security Policy Workshop at George Washington University, and the Center for International Security and Cooperation Workshop at Stanford University.

One of the most formative experiences of the last six years was my time working at Meta. Patrick Lam's care, patience, and dedication as a manager made me a better researcher. Andi Gros, Carlos Velasco, Eugenia Giraudy, Mike Bailey, Pablo Barberá, Paige Maas, Ravi Sojitra, Shankar Iyer, and Taylor Brown all inspired me to do better work, and I am grateful for the time they invested in me.

My parents, Mark and Judy Cuppernull, have always given me limitless support in whatever I decide to pursue in life. From spending nearly every high school and undergraduate weekend in a volleyball gym, to encouraging me when I quit my first "real" job to try to make a living as a musician, to helping me transition to graduate school, they have never wavered. Their commitment to my success made this dissertation possible, and I hope to be as good of a parent one day as they have been to me.

And finally, to Caroline. Thank you for being my biggest supporter and always helping me see the bigger picture. Your love and companionship have made graduate school, and life, immeasurably better.

Abstract

This dissertation makes the case that emerging computational technologies—encompassing machine learning tools, the data that powers them in commercial settings, and the consumer services and devices they optimize—hold theoretical and methodological significance for the field of international relations. These technologies enrich our understanding of international politics in two ways. First, computational technologies hold consequences for understanding the contemporary international order, specifically because technology firms use them to shape digital information transmission. Second, scholars can use these same computational technologies to offer novel insights across diverse debates in the field.

In the first chapter, I argue that large technology companies are critical actors in international politics due to their ownership of data collection and processing infrastructures. In short, digital trace data—the data generated by interactions with internet-connected devices—and the technologies that collect and process it allow technology firms to set the terms by which an immensely differentiated world is commonly understood. I draw on Susan Strange's work on structural power to argue that technology firms shape the choice sets of other actors by mediating information environments. This permits these firms to acquire structural power, which they exercise in at least three ways. First, firms mediate the distribution of legibility derived from digital trace data, whereby they decide the terms under which information is distributed. Second, firms seek to maximize their profit by shaping individual interests, often affecting political outcomes. Finally, firms use structural power to perform state-like functions both domestically and internationally.

The second chapter discusses the role of internet search engines in shaping digital information flows, a subject surprisingly neglected in international relations literature. Co-authored with Dr. Rochelle Terman, this chapter presents an audit of Google search engine result

pages across international affairs topics, revealing three primary findings. First, we find substantial variation in the reach of ideological content, including state propaganda and material from transnational advocacy organizations. Second, search results strongly correlate with search language, suggesting that language is a primary factor mediating exposure to political information. Finally, we analyze search results related to the war in Ukraine generated both before and after the 2022 Russian invasion and find more pronounced geographic clustering in post-invasion results, especially among states in less common language groups.

In the final chapter, I demonstrate the utility of computational technologies in providing novel insights into other debates in the field. I develop a word embedding methodology to quantify friendship and collective identity between states. By applying this methodology to a corpus of all speeches in the United States Congress from 1899-2017, I create biennial friendship and collective identity measures between the United States and 192 countries. I find that perceptions of friendship between the United States and other countries grew dramatically following World War II, while perceptions of collective identity have slowly grown over the past eight decades. Moreover, the correlates of collective identity evolved over the 20th century to match those that predict enmity, suggesting the existence of a shared identity among enemies in international politics.

Chapter 1

Introduction

This dissertation makes the case that emerging computational technologies hold both theoretical and methodological significance for the field of international relations. "Computational technologies" collectively refers to machine learning tools, the data that powers them in commercial settings, and the consumer services and devices these tools are designed to optimize. How do these technologies enrich our understanding of international politics? My contention is twofold. First, as an area of study, computational technologies hold consequences for understanding the contemporary international order, specifically because technology firms use them to shape digital information transmission. Second, as a set of research tools, scholars can use computational technologies to offer novel insights across diverse debates in the field.

Let us first consider computational technologies as an area of study. Every day, interactions with billions of internet-connected devices generate a vast amount of data. This data, known as digital trace data, is predominantly created, stored, and analyzed by private technology firms. The capacity of these firms to comprehend human interests and behavior is unprecedented in its scope, scale, and immediacy, surpassing the capabilities of historical consumer data collection regimes and contemporary state intelligence agencies. What is more, digital trace data often serves the purpose of optimizing the products and services that initially collected it, creating a feedback loop of humans, technology, and firms, where data is both constituted by human experiences and constitutive of them. If current trends are any indication, computational technologies will only continue to become more deeply entrenched in social and political life.

What are the consequences of computational technologies for understanding the contemporary international order? This question holds practical importance. These technologies have profoundly impacted many aspects of politics. From malicious actors leveraging social media to affect democratic elections, to the most recent wave of artificial intelligence technologies standing to disrupt labor markets, understanding the ramifications of computa-

tional technologies is paramount for policymakers and the public as they navigate evolving political landscapes.

This question also carries theoretical importance. Increasingly, individual information seeking occurs online. Because content algorithms filter and transmit diverse information to individuals depending on their geographic location, spoken language, predicted interests, and other factors, technology firms wield the authority to differentially portray international realities across different demographics. Theorizing perception in international politics will only grow more difficult without considering how technology firms use computational technologies to shape information accessibility. Shoshana Zuboff raises a simple, yet crucial question regarding the determination of information access: "who decides?"[1] Technology firms play a pivotal role in deciding the global distribution of information.

Chapter 1 presents my main theoretical contribution on this subject. I argue that the world's largest technology companies are critical actors in international politics due to their ownership of powerful data collection and processing infrastructures. In short, my argument is that digital trace data and the technologies that collect and process it allow technology firms to set the terms by which an immensely differentiated world is commonly understood. I draw on Susan Strange's work on structural power to argue that technology firms—who acquire, store, analyze, and communicate vast amounts of information—shape the choice sets of other actors by virtue of their mediation of information environments. This permits these firms to acquire structural power, which they exercise in at least three ways. First, firms mediate the distribution of legibility derived from digital trace data, whereby they decide to whom, when, and the terms under which information is distributed. Second, firms seek to maximize their own profit by shaping individual interests, often affecting political outcomes. Finally, firms use their structural power to operate akin to a state, performing state-like functions both domestically and internationally. Domestically, technology firms provide

[1] Zuboff 2019, p. 181.

election security, emergency response management, law enforcement, and a communications utility. Internationally, technology firms gather and transmit intelligence and weaponize international information networks to impose costs on states.

Next, I consider computational technologies themselves as research tools. Many of these technologies, including digital trace data and machine learning models, can serve as tools for researchers studying a broad range of questions in international relations. I address two distinct questions in Chapters 2 and 3, demonstrating the efficacy of such tools in studying both computational technologies themselves as well as broader questions in the field.

Chapter 2 focuses on one specific computational technology: internet search engines. Past research shows that Google search results for topics like "Tahrir Square" can vary significantly, where some individuals obtain news related to the Egyptian Revolution while others see links to travel agencies.[2] Co-authored with Dr. Rochelle Terman, this chapter sets out to quantify the extent to which internet search platforms differentially transmit information on political topics between countries. This contributes to the formation of what we term "information pools," which reflect distinct configurations of information available to individuals when they search about a particular topic.

In this chapter, we conduct a large-scale audit of Google search engine result pages across various international affairs topics. We leverage, among other methods, computer vision tools to conduct this audit. Our analysis reveals three primary findings. First, we find substantial variation in the reach of ideological content, including state propaganda and material from transnational advocacy organizations. Second, localized search results strongly correlate with search language, suggesting that language is a primary factor mediating people's exposure to information about international affairs. Finally, we analyze search result pages related to the war in Ukraine generated both before and after the 2022 Russian invasion and find more pronounced geographic clustering in post-invasion results, especially among states in

[2]Pariser 2011.

less common language groups. This chapter sheds light on the role of technology companies in shaping perceptions of international affairs.

In Chapter 3, I demonstrate the utility of computational technologies in providing novel insights on broader, historical debates in the field. I develop a neural network-based word embedding methodology to quantify the perception of friendship and collective identity between states. Despite policymakers and the public often invoking the languages of friendship and collective identity to describe international politics, these concepts have remained difficult to measure and study in the absence of clear formal proxies, especially considering that they are fundamentally rooted in actor perception.

In this chapter, I create biennial friendship and collective identity measures between the United States and 192 countries by applying this word embedding methodology to a corpus of 1.7 billion words taken from all speeches in the United States Congress from 1899-2017. Exploring this data set, I find that perceptions of friendship between the United States and other countries grew dramatically immediately following World War II, while perceptions of collective identity with other states have slowly grown over the past eight decades. Moreover, the correlates of collective identity evolved over the 20th century and have recently come to match those which predict enmity, suggesting the existence of a shared identity among enemies in international politics.

The primary contribution of this dissertation to the field of international relations is in the presentation of computational technologies as both an area of study and a set of research methods. In Chapter 2, I establish a theoretical framework to consider how the control of information facilitated by computational technologies permits technology firms to gain power in the context of international politics. Specifically, this chapter underscores a novel mechanism through which non-state actors wield influence on global politics—via the creation, analysis, and communication of vast amounts of data. Chapter 3 addresses the surprising dearth of research in the field on the role of internet search engines in explaining

heterogeneity in information availability worldwide. This chapter complements previous research on the role of state actors—through activities like digital censorship or propaganda production—in explaining variation in information accessibility. Chapters 2 and 3 collectively address theories of perception and misperception by highlighting the mediating role of technology firms in contemporary information seeking.

Chapter 4 makes a methodological contribution to international relations research by introducing the dimensional approach to word embeddings to the field. This methodology is transferable to the study of virtually any subject in international relations and allows for the quantitative study of concepts which theoretically cannot be reduced to material factors. Moreover, its potential utility is set to expand as more large text corpora become digitized. Substantively, this chapter explicitly analyzes distinct components of geopolitical affinity, a concept somewhat ambiguously approached in prior empirical studies. I disentangle various facets of geopolitical affinity, finding that perceptions of friendship and collective identity diverge from one another and have evolved over time.

Chapter 2

You Have One New Notification: Digital Trace Data and the Role of Technology Firms in International Politics

2.1 Introduction

On the day of the 2010 United States midterm elections, Facebook conducted an experiment involving 61 million users. The randomized control trial included all U.S.-based Facebook users over the age of 18 who accessed the platform on November 2, 2010. Facebook assigned users to one of three groups: 1) the "social message" group, which received a statement in News Feed with information about local polling places, a clickable "I Voted" button, a tally of users who clicked the button, and images of their Facebook friends who had clicked "I Voted"; 2) the "informational message" group, which received all the above information except for the images of friends; or 3) a control group that received no message. The study found that users in the "social message" group were 2% more likely to click the "I Voted" button than those in the "informational message" group. Furthermore, the "social message" group showed a 0.39% increase in election turnout, as verified through matching Facebook users with public voting records. In total, the experiment prompted approximately 230,000 people to vote who would not have otherwise and 12 million people to publicly share that they had voted. Facebook claimed it had "directly influenced political self-expression, information seeking and real-world behavior of millions of people."[1]

Has the emergence of digital trace data (DTD)—the data generated by interactions with devices connected to the internet—afforded technology firms new power in the context of international politics? More generally, what is the role of technology in international relations (IR)? In this chapter, I argue that the unique nature of DTD and the tools that process it allow technology firms to set the terms by which a vastly differentiated world is commonly understood. As a result, technology firms affect the global distribution of legibility, shape individual interests, and perform state-like functions.

I identify three distinguishing characteristics of DTD that make it a historically unique

[1] Bond et al. 2012.

form of information. First, DTD captures data on subject *engagement*—reflecting the active response to a stimulus—whereas previous data enterprises concerning whole populations only collected data on subject *impressions*, representing the passive receipt of a stimulus. Second, DTD is easily duplicatable, meaning it can be copied and transported instantaneously at virtually no cost. While the costs of data duplication and transportation have shrunk over time, it is only with digital data that these costs have lowered to near zero. Finally, DTD provides multidimensional information in real-time. DTD infrastructures continuously collect diverse data points to maintain a current view of the subjects within their architectures. Previous large scale data regimes—including map making, national censuses, and firm market research—only updated periodically. While dragnet phone surveillance of the 20th century updated more regularly, it lacked the nuanced detail of more modern technologies.

Drawing on Susan Strange's concept of structural power, I argue that technology firms, with their acquisition, storage, analysis, and communication of vast amounts of information, shape the choice sets of other actors by mediating information environments. Originally described as a feature of states, Strange defines structural power as the power to shape and determine the structures of the global political economy within which other actors operate.[2] Strange argues that controlling and determining access to knowledge is one pathway through which actors accrue structural power.

In this fashion, I argue that technology firms accumulate structural power by virtue of their collection, storage, analysis, and communication of DTD. Firms exercise structural power in at least three ways. First, firms mediate the distribution of DTD-derived legibility, whereby they control to whom, when, and the terms under which information is distributed. Second, firms seek to maximize their own profit by shaping individual interests. While this often entails maximizing user "screen time" on a particular service, firms have also actively attempted to shape the political behavior of individuals. Finally, technology firms use their

[2]Strange 1988, pp. 24–25.

structural power to perform state-like functions, both domestically and internationally. Domestically, they provide services such as election security, emergency response management, law enforcement, and communication utilities. Internationally, they gather and transmit intelligence and weaponize international information networks to impose costs on states. I do not claim this list of functions to be exhaustive, but rather illustrative of the various ways in which the control of knowledge enables technology firms to operate akin to a state.

This analysis adds to several literatures. I address the role of technology in IR by emphasizing the endogenous nature of DTD. This challenges the prevailing notion in the field that treats DTD as an exogenous form of information, neglecting the crucial role played by various actors, particularly technology firms, in generating and shaping data. By highlighting the constitutive role of these actors in the production of data, my analysis seeks to provide a more comprehensive understanding of the interplay between technology and international politics. Second, this research engages with the literature on perception and misperception in IR. By acknowledging that online platforms mediated by large technology firms shape contemporary information seeking, this chapter highlights the subjectivity and non-neutrality of these platforms, which provide different information to different users.[3] Lastly, this research expands upon non-state actor theories in IR by highlighting an additional mechanism—the exploitation of vast amount of data—by which non-state actors can influence international politics.

The rest of this chapter proceeds in six parts. I begin by defining DTD and examining its present scope and scale. Part 3 evaluates the treatment of technology in different strands of research in IR before exploring the existing literature on DTD. Part 4 introduces Strange's concept of structural power and demonstrates its relevance in examining DTD, while Part 5 discusses several pathways through which technology firms exercise structural power in political contexts. Part 6 connects this theory to broader debates in IR and highlights the

[3]Luo, Puett, and Smith 2023; Ochigame and Ye 2021.

utility of the structural perspective. Part 7 concludes.

2.2 Digital Trace Data

2.2.1 What is Digital Trace Data?

Digital trace data refers to the "records of activity (trace data) undertaken through an online information system (thus, digital)."[4] Effectively, DTD is the data created by interactions with devices connected to the internet. DTD exists in countless forms—records of likes and comments on various social media platforms, "offline" conversations recorded by smart devices, video logs recorded by home smart cameras, user location data collected by smartphone applications, or driving data collected by WiFi-enabled consumer vehicles. The vast majority of DTD is proprietary, whereby technology firms use the DTD they collect to optimize the same products and services performing the collection. DTD has extensive applications across industries, informs policy-making, and plays a crucial role in academic research. It also is central to information-seeking, with 50% of Americans "often" or "sometimes" receiving news via social media as of 2022.[5]

2.2.2 How Much Digital Trace Data Exists?

Technology firms collect "too much data."[6] Despite individuals' strong preference for privacy, virtually all internet users reveal their personal data. Economics literature explains this puzzle as a function of data externalities.[7] A data externality exists when data about one individual reveals or updates the probability of data about another individual. When users share their data with technology firms, they inadvertently expose the data of corre-

[4]Howison et al. 2011, p. 769.
[5]Liedke and Matsa 2022.
[6]Acemoglu et al. 2019.
[7]Acemoglu et al. 2019; Bergemann, Bonatti, and Gan 2021.

lated individuals, thereby diminishing the value of others' data. As a result, individuals tend to undervalue their own data and engage in excessive data sharing, leading to the "overcollection" of data by firms.

The sheer volume of DTD is staggering. While precise estimates are challenging to create, IBM estimated that a total of 40 trillion gigabytes of data would be generated globally in 2020 alone.[8] Out of this amount, 15% would be original data, equating to approximately 773,000 gigabytes of original data per person worldwide during the year. To put this amount into perspective, it is equivalent to the storage capacity of 3,000 baseline 2023 MacBook Pros.[9] The proliferation of smart devices has contributed significantly to the surge in data production. There are approximately 14.3 billion internet of things (IoT) endpoints worldwide, which include in-home smart devices, wearables, cameras and sensors, and self-driving vehicles.[10] Internet usage continues to grow in developing countries, with Meta alone attracting 3.81 billion monthly active users across their suite of products globally in 2023.[11] In summary, technology firms accumulate vast amounts of data.

2.2.3 Who Collects Digital Trace Data?

Data markets tend to exhibit oligopolistic characteristics due to the substantial fixed costs associated with establishing the necessary infrastructures for collecting and processing DTD.[12] Only the largest firms have the financial capacity to bear these costs, and social networks with larger user bases are more appealing to consumers.[13] Some firms offer services that translate fixed costs to variable costs, which further allows a small number of firms to achieve even greater economies of scale—for example, Amazon Web Services provides

[8] Dailey 2020.
[9] The baseline MacBook Pro offers 256GB of storage.
[10] Sinha 2023.
[11] *Meta Platforms, Inc. Form 10-Q* 2023.
[12] Bessen 2020.
[13] Carrière-Swallow and Haksar 2019.

computing power on a variable scale, eliminating the need for organizations to build their own hardware facilities. Consequently, it is not surprising that in terms of market capitalization, the top data firms are among the largest companies in the world—as of December 31, 2022, Apple, Microsoft, Alphabet (Google), and Amazon were the four largest publicly traded American firms. The combined market capitalization of these four companies was $5.85 trillion, larger than the GDP of all countries other than the U.S. and China.

These firms dominate their respective market sectors:

- In online search, Google maintains over 90% market share. It is so popular that the most popular search term on Bing.com is "Google".[14]
- In e-commerce, Amazon controls over half of the global market share in online department store retail, with revenue in 2020 reaching $277 billion, more than seven times the revenue of its closest U.S. competitor, Walmart.[15]
- In social networking, Meta operates four of the top five platforms in terms of monthly active users: Facebook (2.9B MAU), WhatsApp (2B MAU), Messenger (1.3B MAU), and Instagram (1.3B MAU).[16]

Furthermore, Amazon, Google, and Microsoft operate the majority of the world's "hyperscale" data centers, each housing at least 5,000 servers.[17] These data centers are primarily located in the United States, with 39% of the servers situated there, followed by 10% in China and 6% in Japan.

[14]White and Bodini 2021.
[15]Trefis 2020.
[16]YouTube attracts 2.3B monthly active users. Figures current as of October 2021.
[17]"Microsoft, Amazon and Google Account for Over Half of Today's 600 Hyperscale Data Centers" 2021.

2.2.4 What Makes Digital Trace Data Unique?

Digital trace data possesses three distinctive characteristics that set it apart from previous forms of data. First, DTD captures goes beyond capturing impressions and includes data on engagement. *Impression* data captures the volume of traffic of a particular item, such as the number of books sold, the wattage of electricity used, the quantity of a good transported via a railroad, or the number of viewers tuned into a television program. In contrast, *engagement* data captures agent response: social media firms record a user's likes and comments, eye tracking technology determines which advertisement on a screen an individual is viewing, and smart devices record "offline" conversations. Where past technologies primarily gathered data on the passive receipt of stimuli by individuals, DTD allows for a deeper understanding of their active engagement. Engagement data permits a detailed form of legibility that was unattainable with previous forms of consumer data. Although it was possible to collect engagement data in the past through methods like market research surveys, DTD is the first technology to enable the collection of such information at scale.

Second, DTD is an easily duplicatable input of production. It is easily duplicatable in the sense that it can be copied and transported instantaneously at virtually no cost. In contrast, prior to the digital era, duplicating data involved significant expenses. While advancements such as the printing press and fax machine reduced the costs of data duplication, it was not until digital data emerged that these costs lowered to near zero. DTD serves as a critical input for various services, ranging from social media platforms to rideshare applications and self-driving cars.[18] The Department of Defense has even likened DTD to "the 21st century equivalent of a global natural resource, like timber, iron, or oil previously."[19] However, unlike these conventional resources, DTD is easily duplicated and transported.[20]

[18]Carrière-Swallow and Haksar 2019.

[19]Trugeur 2019.

[20]DTD is also unique as an input of production because it is nonrival. Unlike timber, iron, or oil, where use by one agent precludes simultaneous use by another agent, the same digital trace data can be used by

A final distinguishing feature of DTD is its high-dimensional nature and real-time updates. DTD infrastructures continuously collect high-dimensional data—data characterized by a large number of variables per observation—to maintain a current view of subjects. Moreover, the machine learning systems that analyze this data continuously improve as they incorporate more information. In contrast, previous data regimes typically allowed for data collection in "snapshots." Real-time data collection methods were either limited to capturing high-level metadata, as seen in dragnet telephone surveillance, or were prohibitively expensive to conduct at scale, such as Cold War-era espionage operations. The high-dimensional and real-time characteristics of DTD present a significant departure from these earlier data regimes.

Table 1 provides a comparison between DTD and various other types of data.

Table 2.1: Data Type Comparisons

Data	Type	Duplication Costs[21]	Update Frequency	Dimensionality
National Census	Impression	High	10 Years	Medium
Market Research	Engagement	High	Per Cycle	Medium
Dragnet Telephone Surveillance	Impression	High	Real-time	Low
Digital Trace Data	Engagement	Very Low	Real-time	High

2.3 Firms, Technology, and Digital Trace Data in International Relations

Modernist writers of the 1960s and 1970s argued that the world was undergoing a fundamental transition, characterized by the emergence of a "global village" brought about by economic interdependence. According to the modernists, this interconnectedness would

multiple actors simultaneously.

render traditional power politics obsolete. They contended that the dominant paradigm of political realism simply did not square with the rising international society of the time. Instead, they believed that multinational corporations, among other actors, would increasingly gain power in the international system and regulate interdependence between states.[22]

Later research critiqued the modernist approach for underestimating the resiliency of the state and the extent to which states could "[exercise] power to shape or distort patterns of societal interdependence."[23] Critics pointed out that even the most powerful corporations have limited sway over the state, and the unique power of multinational corporations did not pose a significant challenge to state authority.[24] The paradigmatic neoliberal institutionalist arguments that followed were founded upon realist principles and offered state-centric narratives regarding interdependence in the international system.[25]

The modernist discourse predominantly approached technology from an instrumental perspective. This literature sought "to understand how world politics *was being affected by* rapid technological change" or how a state's technological advantage mediates the effect of global economic openness.[26] Other important works in the field have similarly tended to treat technology as exogenous, considering it as an existing tool used by states to achieve their goals. The offense-defense balance literature, for instance, posits that the characteristics of military technology during a particular time have deterministic effects in the international system.[27] Several classic works on nuclear technology discuss the distinct effects of nuclear weapons on international stability, with two strands of literature arguing that nuclear weapons have and have not changed international politics.[28] Waltz (1979) writes that state power is a function of a state's "size of population and territory, resource endowment, economic capability, mil-

[22]Angell 1969; S. Brown 1974; L. R. Brown 1972; Kindleberger 1969.
[23]Keohane and Nye 2012.
[24]Krasner 1978; Ruggie 2018.
[25]Keohane 1984.
[26]Krasner 1976; Keohane and Nye 2012, xxiv, emphasis added.
[27]Glaser 2010; Glaser 1996.
[28]Jervis 1990; Lieber and Press 2017; Schelling 1966; Tannenwald 1999.

itary strength, political stability and competence", and later argues that the United States can "substitute for critical materials" by virtue of their superior technology—in essence, claiming that technology can substitute for an exogenous resource base.[29] The United States and the Soviet Union were economically superior to other states, which allowed them to "exploit military technology on a large scale and at the scientific frontiers."[30] In other words, by virtue of their economic strength, they could better use technology than other states. Finally, Keohane (1984) briefly touched upon the notion that "technological advances", in addition to domestic institutions, basic resources, and military power, had contributed to American hegemony.[31] However, this idea was not further developed, as Keohane treated these variables as "background conditions" for hegemony, essentially considering them as exogenous state attributes that influenced the likelihood of hegemonic power.

Not all works, however, have approached technology solely from an instrumental perspective. Other research has addressed the endogenous nature of technology. Wendt (1999) argues that technology "is created by purposeful agents and embodies the state of their technical knowledge (ideas) at that time."[32] He further claims that "technological artifacts do have intrinsic causal powers", and that a stripped-down view of technological determinism is compatible with his view of social constructivism. Jervis (1976) parenthetically discusses the co-constitution of technology and actors. He contends that states develop distinct weapons systems by virtue of "their own experiences", such as the efficacy of certain weapons in past wars.[33] Jervis also notes that technological conditions can alter actor perception to increase the probability of war.[34] From this perspective, Jervis argues that experience constitutes technology, and technology constitutes experience. In summary, while some works in IR

[29]Waltz 1979, pp. 131, 146.
[30]Ibid., p. 181.
[31]Keohane 1984, p. 41.
[32]Wendt 1999, p. 111.
[33]Jervis 1976, p. 242.
[34]Ibid., p. 67.

have explored the endogenous aspects of technology, the majority of paradigmatic works treat technology as exogenous.

IR scholars have largely treated DTD through an instrumental lens, approaching the implications of DTD in two primary ways.[35] The first involves the transformation of intelligence practices. Since Edward Snowden's revelations about state surveillance programs, there has been significant public and scholarly interest in the use of digital data by intelligence agencies.[36] Undoubtedly, the vast amount of data on individuals worldwide serves as a crucial source of intelligence. Additionally, this data has opened various forums for interstate cooperation, for example via data-sharing programs within the Five Eyes intelligence alliance (comprising the U.S., U.K., Canada, Australia, and New Zealand).[37] IR scholars have also explored the impact of DTD in the context of social media.[38] There is a rich literature on the proliferation of fake news, in which various studies have shown that fake news travels quickly, erodes public trust in institutions, and opens pathways for states to take coercive action against each other.[39] Social media has also been linked to extremist recruitment, public disinformation campaigns, and political polarization.[40]

These studies predominantly analyze power dynamics in relation to DTD through regulative and instrumental perspectives, focusing on how actors employ DTD and associated tools to achieve specific objectives. In some cases, there is a risk of adopting a technologically determinist viewpoint, treating DTD as an external force that independently shapes outcomes. However, it is essential to recognize that DTD is not solely a technological tool but also a valuable source of information about human behavior. By design, it is inherently

[35] A notable exception is (S. Srivastava 2021), who argues that the algorithmic governance in which large technology firms participate has created new forms of private authority.

[36] Bauman et al. 2014; Greenwald 2014; Lyon 2015; Preibusch 2015.

[37] Bigo and Bonelli 2019; *Promoting Public Safety, Privacy, and the Rule of Law Around the World: The Purpose and Impact of the CLOUD Act* 2019; Trugeur 2019.

[38] Sarah E Kreps 2020.

[39] Ognyanova et al. 2020; Verma, Fleischmann, and Koltai 2017; Vosoughi, Roy, and Aral 2018; Ziegler 2018.

[40] Bail 2021; Bradshaw and Howard 2018; Geist and Lohn 2018; Mitts 2019.

embedded within social life and thus fundamentally endogenous to it.

A volume of work spanning the social sciences has argued for the co-constitution of information technologies and social life. Scott (1998) argues that the means by which a state sees its population has a constitutive effect on the population itself.[41] Anderson (1993) discusses how maps, censuses, and museums helped to consolidate the national identity of colonial populations—once colonizers developed systems to make colonies more legible, the colonized populations began perceiving themselves in accordance with these schemes.[42] Branch (2011) posits that cartographic technology in the early modern era changed how actors thought about political space and organization, ultimately shaping the formation of the sovereign state. Specifically, he argues that "changes in representational technologies structured political interactions, but only because those technologies altered ideas about the appropriate and legitimate forms of political authority."[43] Isin and Ruppert (2019) contend that the British census in colonial territories formed a type of data politics which rendered colonial populations as objects of power. The census categorized populations into predefined groups, not primarily to provide a more accurate description of the population, but rather as a means of governing the empire in a different manner.[44]

DTD, as an information source, is intrinsically tied to human behavior. It is not only the technologies developed by profit-driven firms that generate, collect, and analyze DTD, but the data itself is a direct result of human actions and interactions. The services and products that are powered by this data, such as social networking sites, online search platforms, and home smart devices, play a significant role in shaping the informational environments that influence actor behavior and interests. These services and devices mold the flow of information and the choices available to individuals, thereby shaping the information environments

[41] Scott 2008.
[42] B. R. O. Anderson 2016.
[43] Branch 2011, p. 3.
[44] Isin and Ruppert 2019, p. 215.

which constitute actor perceptions, preferences, and actions.

2.4 Digital Trace Data and Structural Power

The framework I present for analyzing the role of technology firms and DTD in international politics is rooted in Susan Strange's work on structural power. Strange (1988) distinguishes between relational power and structural power.[45] Relational power refers to the ability of Actor A to compel Actor B to take an action they would not have otherwise taken. Within IR, much of the analysis of DTD has focused on how it grants relational power to actors, such as enhancing the capabilities of intelligence agencies and law enforcement or enabling foreign actors to manipulate democratic elections through social media platforms. On the other hand, structural power "confers the power to decide how things shall be done [and] the power to shape frameworks within which [actors] relate to each other."[46] Those who possess structural power can influence the decision-making environment of other actors without directly coercing them into specific choices.

According to Strange, there are four pathways by which an actor can accumulate structural power. First, structural power comes from control over the security of other actors. Second, it is conferred to those who determine the manner and mode of production of essential goods and services. Third, structural power resides with those who determine access to credit, particularly within advanced economies. Fourth, and most importantly of the purposes of this chapter, structural power can be "exercised by those who possess knowledge [and] can wholly or partially limit or decide the terms of access to it."[47]

Strange argues that structural power is conferred to those who are acknowledged by society to possess desirable knowledge, are engaged in its acquisition, entrusted with its storage,

[45]Strange 1988.
[46]Ibid., p. 25.
[47]Ibid., p. 30.

and control the channels by which it is communicated. She emphasizes that structural power is accumulated by actors at key decision-making positions in knowledge structures. Knowledge structures determine "what knowledge is discovered, how it is stored, and who communicates it by what means to whom on what terms."[48] In the realm of DTD, a small number of oligopolistic technology firms hold these pivotal decision-making positions within knowledge structures.

2.4.1 What Knowledge is Discovered

Technology companies collect and analyze vast amounts of raw data that is used for diverse purposes, including targeted advertising, optimizing autonomous vehicles, and complying with data requests from state intelligence agencies. However, the term "raw data" is a misnomer. Data is never "raw"—an actor selectively chooses to create some pieces of data over others. Technology firms actively choose which data to create and ultimately determine what counts as data. The process of "datafication," which involves transforming human activity into data, is inherently a social process.[49] In addition to data collection, firms also determine the process by which data is transformed into knowledge. For example, Meta employs thousands of product managers to decide how data ought to be used. Furthermore, the company maintains an "industry-leading research program" consisting of hundreds of Ph.D. researchers. This program aims to understand the firm's societal impact and develop cutting-edge artificial intelligence technologies.[50]

2.4.2 How Knowledge is Stored

Firms determine the way in which data is stored. They determine the geographic locations where data is stored through their operation of the majority of hyperscale data centers

[48] Strange 1988, p. 117.
[49] Flyverbom and Murray 2018.
[50] Zuckerberg 2021.

worldwide. The geographic distribution of data confers power to certain actors, as states with privileged positions in information networks can conduct surveillance on other actors and restrict their access to information.[51] Of course, firms have the authority to decide how their databases are structured and the conditions under which knowledge is stored. For example, Google deletes user history records older than 18 months for newly created accounts (existing users can choose to opt into this feature), while Meta retains data for as long as it is necessary to provide products and services to users.[52]

2.4.3 Who Communicates Knowledge by What Means, to Whom, and on What Terms

Technology firms selectively disclose the data and knowledge they accumulate. Meta, for example, received 239,388 total government requests for user data from July to December 2022, of which it produced data for 76.8%.[53] However, the rate of compliance varied significantly depending on the requesting state. Among countries that made at least 100 requests, Meta fulfilled requests at the highest rate for Finland (90%), the United States (88%), and Taiwan (87%), while it had a 0% compliance rate for Hong Kong. This suggests a potential inclination by the firm to disclose data to more liberal states.[54] In response to growing concerns regarding the impact of social media on mental health, Instagram introduced a feature that hides the "like" counts of images on the platform. This feature was meant to "'depressurize people's experience' on the platform," and has since become an optional control for users to enable.[55] Meta also allows users to have control over the extent to which their data is publicly shared. For example, during the withdrawal of U.S. military forces from Afghanistan in 2021, Meta introduced an option for Afghan users to hide their

[51] H. Farrell and Newman 2019.
[52] *Data Policy* 2016.
[53] Meta 2023b.
[54] Ibid.
[55] Criddle 2021.

public information.[56] Additionally, Meta has made various data sets and AI tools publicly available to researchers, including their "Social Connectedness" geographic data set and the RoBERTa model for natural language processing.[57]

2.5 Consequences: How Technology Firms Exercise Structural Power

I identify three mechanisms through which technology firms exercise structural power. First, they control the degree of legibility afforded to other actors. Second, as a means to increase profits, technology firms seek to maximize user engagement. This is achieved through various means of "behavior modification", which can influence and shape individual interests and behaviors. Third, technology firms wield structural power by performing functions that are typically associated with states, often operating in tandem with states themselves.

2.5.1 The Distribution of Legibility

James Scott's seminal work *Seeing Like A State* refers to legibility as the modern state's attempt to "arrange the population in ways that simplified the classic state functions of taxation, conscription, and the prevention of rebellion."[58] More recent work on legibility has expanded its scope to include "the breadth and depth of the state's knowledge about its citizens and their activities", and explored other themes including the role of non-state leaders in facilitating state legibility.[59] Technology firms today determine when, under what terms, and who gets access to the unique lens of legibility that is afforded by high-dimensional real-time engagement data. Technology firms determine when legibility is provided. Meta and

[56] Mozur 2021.
[57] *Facebook Data For Good Social Connectedness Index* 2021; *RoBERTa* 2021.
[58] Scott 2008, p. 2.
[59] Lee and N. Zhang 2017.

Twitter enforce minimum age requirements for users of their services: Facebook, Instagram, and Twitter require users to be at least 13 years old to create an account.[60] Furthermore, technology firms have the capacity to restrict legibility in response to specific events. For instance, Meta removed news content in Australia and concealed posts from numerous government pages following a dispute with the Australian government over publisher payments for online news.[61] Google also threatened to make Google Search unavailable in Australia during the same dispute.[62]

Firms determine the terms under which legibility is provided. Google, Meta, Amazon, and countless other firms optimize content streams to align with their revenue maximization goals, often prioritizing paid advertisements and content that is predicted to be relevant to individual users. However, their methods of designing content feeds have varied. For instance, Google introduced personalized search results for Google Account holders as a part of its main search algorithm in 2005, seven years after the company's incorporation.[63] Personalized search was expanded to all users, regardless of account login, in 2009.[64] Google also piloted "Social Search" in 2009, incorporating a user's social circle's activity into the search results, which was later integrated into the main algorithm in 2011.[65] Initially, Google allowed users to switch between their personalized results and the main results, but this functionality has been removed. Currently, Google operates on "semantic search," which predicts the user's intended query and provides results based on that prediction rather than the explicit query itself.[66] As a result, what was once portrayed as a neutral search platform offering uniform access to information has transformed into a highly personalized and subjective tool that presents different information to different users. In essence, legibility is selectively provided

[60]Facebook 2021.
[61]Isaac et al. 2021; Kaye n.d.
[62]Cave 2021.
[63]Kamvar 2005.
[64]Horling and Kulick 2009.
[65]Heymans and Viswanathan n.d.
[66]Nayak 2019.

based on the interests of advertisers, the predicted interests of users, and other relevant factors.

In a similar vein, social media platforms largely generate revenue by maximizing user screen time, leading them to optimize content and notifications accordingly. Content streams vary drastically between users based upon their predicted interests, geographic location, social circle, and other factors. Chris Bail describes social media as a "prism"—not a mirror of everyday life, but a refraction of it that serves to maximize user engagement, often by amplifying extreme voices.[67]

Technology firms determine who gains access to legibility. At the state level, DTD is a crucial source of intelligence for state intelligence agencies, with a growing trend of outsourcing intelligence gathering to private entities in recent years.[68] On an individual level, several technology firms are actively involved in infrastructural projects aimed at expanding internet access. Meta, for instance, has undertaken the construction of internet infrastructure in developing countries and is currently engaged in laying thousands of miles of underwater sea cables, an endeavor that will provide internet access to hundreds of millions of people.[69] These projects have wide-ranging impacts on various aspects of society, including labor needs and public health:

> "We've seen first-hand the positive impact that increased connectivity has on communities, from education to healthcare. We know that economies flourish when there is widely accessible internet for businesses."[70] – Najam Ahmad, Vice President, Network Infrastructure at Facebook

Notably, Meta, Google, and other firms have partnered together on the Apricot subsea cable project, scheduled for completion in 2024, which "will deliver much-needed internet

[67] Bail 2021.
[68] Trugeur 2019.
[69] Ahmad and Salvadori 2020.
[70] *About | 2Africacable.Com* 2021.

capacity, redundancy, and reliability to expand connections in the Asia-Pacific region."[71]

In summary, technology firms play a significant role in the global distribution of legibility by mediating information flows. Existing research has demonstrated the pivotal role of internet access and digital information as key factors influencing political participation, international trade, and political learning.[72]

2.5.2 Structuring of Individual Interests

DTD is both constituted by and constitutive of subjects. On one hand, DTD is constituted by subjects through their everyday behaviors and interactions. Activities such as posting on a social media platform, walking through a park, or speaking within earshot of an Amazon Echo device all permit the creation of DTD. In effect, technology firms commoditize human experiences and actions.[73]

Crucially, though, DTD also is *constitutive of subjects*. Technology firms seek to influence user actions in various ways to maximize their profit. Examples of this include:

- Mapping technologies, such as Google Maps, which guide individuals along optimized routes to work or the airport or direct them to the nearest coffee shop. In exchange for these services, firms collect user location data, enabling them to understand users' habits and preferences and sell targeted advertising opportunities to third parties.

- Augmented reality games like Pokémon Go leverage DTD to incentivize users to explore specific locations. Launched in 2016, Pokémon Go users physically searched around their local communities for in-game incentives that were geographically positioned through the application's Google Maps integration. Through in-game "micro-

[71] Roehrich 2021.

[72] Campante, Durante, and Sobbrio 2018; Dimitrova et al. 2014; Freund and Weinhold 2004; Gil de Zúñiga, Molyneux, and Zheng 2014; S.-H. Kim 2008; Placek 2020; Tang and Huhe 2014.

[73] Zuboff 2019.

transactions", businesses could incentivize users to visit a particular location, for instance by paying to place a virtual Pokémon inside the bathroom stall of a restaurant.[74]

- Wearable devices use prompts and reminders to encourage users to engage in healthy behaviors, such as meeting daily step goals or maintaining consistent sleep patterns.

Firms also engage in the modifying political behavior. A notable example is Facebook's 61-million user experiment, which generated hundreds of thousands of votes and prompted millions of users into acts of political self-expression (clicking the "I Voted" button).[75] This experiment directly demonstrates the structural power of technology firms. In this case, Facebook did not coerce or explicitly encourage individuals to take specific actions but rather influenced user interests by shaping their informational environment.

Furthermore, former Google Chairman Eric Schmidt personally assisted the 2012 Obama presidential campaign in recruiting personnel and designing its digital strategy. The campaign's data analytics team aimed to understand the preferences of individual voters and employ targeted digital or personal contact to persuade them, as Schmidt explained, using "scientific techniques to predict how people will behave when faced with choices or questions."[76] During the 2021 India state elections, Meta created election-day reminders and other informational content on Facebook specifically designed for easy sharing among users. Extensive research indicates that person-to-person voter mobilization efforts are particularly effective in increasing voter turnout.[77]

Social networking sites also foster "audience effects" in relation to participation in social causes.[78] An audience effect "arises when a person's behavior changes because they believe

[74]Zuboff 2019.
[75]Bond et al. 2012.
[76]Joshua Green 2013.
[77]Gerber, D. P. Green, and Larimer 2008; *Our Steps to Protect State Elections in India* 2021; Schein et al. 2021.
[78]Hamilton and Lind 2016.

someone else is watching them."[79] Social networking sites amplify the visibility of political behavior and support for social causes, which can influence individuals to modify their own behavior. For instance, the 2020 #BlackOutTuesday trend, which saw participation from 28 million Instagram users, faced criticism for being "largely performative" and "catered to the people who want to show that they care."[80] Research focusing on fitness-related social networking platforms reveals that the amount of exercise an individual completes and subsequently shares with their network is positively linked to the size and activity of their online social network, holding other factors constant.[81] The audience effects created by social networks alter the incentives individuals have to engage in specific activities.

2.5.3 Performing State-like Functions

Meta CEO Mark Zuckerberg has gone as far as to claim that "Facebook is more like a government than a traditional company."[82] The structural power firms accrue by mediating information flows permits them to perform a variety of state-like functions. Within domestic contexts, firms provide election security, emergency response management, law enforcement, and communication utilities. At the international level, firms collect and transmit intelligence and weaponize global information networks to impose costs on states.

Election Security

Firms including Google, Meta, and Microsoft are heavily involved in election security worldwide, a task traditionally handled by governments. Google, for instance, offers its *Protect Your Election* suite of products, which provides tailored digital security tools for various stakeholders involved in elections, including campaign managers, candidates, election and

[79] Hamilton and Lind 2016.
[80] Bludau 2021; Jennings 2020, Emphasis added.
[81] *A Big Data Approach to Assessing the Impact of Social Norms: Reporting One's Exercise to a Social Media Audience - Christopher J. Carpenter, Chandra S. Amaravadi, 2019* 2021.
[82] Foer 2017.

news website operators, and journalists. These tools aim to safeguard individuals from phishing attacks, protect websites from DDoS attacks, and automate content moderation to enable discussion websites to "remain open as a positive forum for political conversation."[83] Meta has also taken measures to combat misinformation and foreign funding of political ads in the lead-up to elections worldwide. The company acknowledged the issue of coordinated misinformation campaigns and "bot" activities, and has stated its commitment to sharing relevant information with government, law enforcement, and other groups to "help aid their investigations and crack down on bad actors."[84] Meta is "actively engaged" with the Department of Homeland Security, the FBI, and other agencies to aid in identifying and disrupting nefarious information operations. Meta claims that it and government agencies have access to unique pools of information, and that by working together they can more effectively identify "coordinated inauthentic behavior."[85]

Microsoft's election verification tool, *ElectionGuard*, aims to "enable end-to-end verification of elections, open results to third-party organizations for secure validation, and allow individual voters to confirm their votes were correctly counted."[86] The firm is working with manufacturers and vendors to incorporate the technology into future U.S. elections. Furthermore, Microsoft offers *Microsoft 365 for Campaigns*, which provides a cost-effective infrastructure for campaigns to enhance their cybersecurity defenses against various threats, including those originating from nation-states.[87] The firm also offers free cybersecurity training for individuals and groups working on elections and a browser plugin that aids users in verifying the credibility of a particular source.[88]

[83] *Google | Protect Your Election | Protection from Digital Attacks* 2021.
[84] Gleicher 2018.
[85] Ibid.
[86] Burt 2019.
[87] Neutze 2019.
[88] Frank 2019.

Emergency Response Management

Firms, in collaboration with governments, actively contribute to emergency response management. Meta's Disaster Maps use geolocated user data to provide valuable insights during natural disasters. These maps help predict evacuation routes, assess cell network connectivity, determine access to electricity, and track long-term displacement.[89] The data used in Disaster Maps includes user location data, information on cell site connectivity, and data on phone battery charging.[90] Meta designed Disaster Maps in conjunction with several humanitarian organizations, including the International Federation of Red Cross and Red Crescent Societies (IFRC), UNICEF, and the World Food Programme. These organizations rely on both private firms and governments for critical resources. For instance, government donations cover 82% of the IFRC's annual budget, demonstrating the collaboration between public and private entities in addressing emergency response needs.[91]

During the COVID-19 pandemic, Google created several initiatives akin to state public assistance programs and provided real-time data and a suite of curated software tools to governments. The firm provided bailout relief to small businesses by pledging $200 million in small business loans.[92] Additionally, Google committed $100 million in direct cash transfers to individuals affected by COVID-19 and provided full-time pro bono work of 10 employees for six months to the State of New York's "COVID-19 Technology SWAT Team."[93] In terms of technological support, Google developed an application for the regional government of Madrid, enabling citizens to perform self-assessments of COVID-19 symptoms.[94] The firm also created a public "Community Mobility Report" in response to calls from public health

[89] Maas 2019.
[90] Ibid.
[91] *The ICRC's Funding and Spending* 2016.
[92] Pichai 2020.
[93] *COVID-19 Response | Google.Org* 2021.
[94] *How Google Cloud Is Helping during Coronavirus* 2021.

officials for COVID-19 data presented in a format similar to Google Maps.[95]

Law Enforcement

In the aftermath of the January 6, 2021 attacks on the U.S. Capitol, the FBI sought assistance from private technology companies to help identify suspects involved in the incident. Leveraging the power of social media, the FBI used platforms like Facebook to crowdsource the identification of individuals who were still at large months after the attacks. The FBI shared CCTV images on its Facebook page and encouraged users to submit tips to aid in the identification process.[96] Meta actively cooperated with ongoing investigations, preserving and disclosing account data of suspects to local and federal law enforcement agencies.[97] In the wake of the events, 45 federal criminal cases referred to Google geolocation data as evidence. This data was obtained through the use of "geofence warrants" served to Google, allowing access to geolocation information that provides precise location data of a user's cellular device within a 10-meter range.[98]

Meta regularly provides data to law enforcement agencies around the world, which has assisted in various criminal investigations. For instance, Meta's data has been used to locate individuals suspected of being foreign terrorist fighters in Italy, apprehend an interstate gang accused of kidnapping and extorting a foreign traveler in India, and locate a fugitive on trial for murder in Brazil who had made death threats against witnesses in their case.[99] In 2020, Meta paid a third-party cybersecurity firm to develop a "hacking tool" for Tails, a privacy-focused operating system that obscures a user's IP address. This tool was later shared with the FBI to aid in the apprehension of a child predator who had been active on the platform.[100] Additionally, DNA testing company FamilyTreeDNA disclosed DNA data from

[95] *COVID-19 Community Mobility Report* 2021.
[96] Investigation 2021.
[97] Brewster 2021.
[98] M. Harris 2021.
[99] Meta 2023a.
[100] Franceschi-Bicchierai 2020.

2 million clients to the FBI, which aided in the resolution of violent crimes.[101] Parler even aided law enforcement, providing account data to the FBI in response to an exigent request involving a user who had made threats to kill Donald Trump and several U.S. Supreme Court justices.[102]

Communication Utilities

Large technology companies provide services central to society and economies of scale have led to oligopolistic concentration. Some legal scholars have advocated for treating companies with these characteristics as public utilities due to the essential nature of the goods and services they provide.[103] Justice Lewis Brandeis argued that when firms offer goods or services that are crucial or exceptionally significant to public life, and market conditions such as economies of scale or other factors impede competition, regulatory measures should be implemented to treat those firms as public utilities.[104]

Technology companies have effectively become communication utilities due to the consolidation of markets and the role they play in providing indispensable services to society. The indispensability of these services was prominently highlighted during the October 2021 Facebook outage, which rendered Facebook and its affiliated platforms inaccessible on a global scale for approximately six hours. Meta's "Free Basics" program provides basic internet access to over 1 billion people worldwide, a service which was unavailable during the outage. The outage also affected more than 50 million businesses that rely on Facebook Messenger or WhatsApp for e-commerce transactions, disrupting their ability to conduct business.[105] The widespread reliance on these services was evident as some communities mistakenly believed that the "entire internet was offline."[106] Much like governments build and maintain

[101] Haag 2019.
[102] Hall 2021.
[103] Rahman 2018; Minow 2021, p. 120.
[104] Brandeis 1932.
[105] Sebastian 2020; Sweney 2021.
[106] Asher-Schapiro and Teixeira 2021.

infrastructure for transportation, sanitation, and other services essential to society, large technology firms provide communications infrastructure that populations—especially those in the Global South—depend on for their daily lives.

Intelligence Services

The U.S. Department of Defense recently claimed that "data must be regarded as one of the most powerful resources in the Department's arsenal."[107] The DTD that technology firms collect is a crucial source of intelligence for nation-states. Through various data sharing partnerships, technology firms have become increasingly integrated into the fabric of national intelligence services. For example, the number of user accounts for which governments made data requests to Google saw a tenfold increase between 2010 and 2020.[108] The intersection of public intelligence agencies and private technology companies is evident in the movement of officials between these sectors. During the Obama administration, 25 national security and intelligence officials joined Google, while three Google executives transitioned to the Department of Defense.[109] This "political-economic fraternity" is born out of the compatible needs of public intelligence agencies and private technology companies—intelligence agencies seek to maximize legibility and firms seek to maximize profit.[110]

Examples of this fraternity abound. Microsoft, Yahoo!, Google, Facebook, and Apple were all involved in the NSA's PRISM program, which was exposed through Edward Snowden's leaks in 2013.[111] Within PRISM, firms were directed to "erect a locked mailbox [for user data] and give the government the key."[112] Private technology companies control critical points within communication infrastructures, and governments maintain regular communi-

[107] Trugeur 2019.
[108] *Requests for User Information – Google Transparency Report* 2021.
[109] *Google's Revolving Door (US)* 2016.
[110] Zuboff 2019.
[111] Johnson 2021.
[112] Miller 2021.

cation with them to obtain necessary data.[113] This is particularly true in the United States, where a significant portion of global internet communications pass through U.S.-based systems, granting U.S.-based intelligence services advantageous access to data held by these firms.[114] The 2018 Clarifying Lawful Overseas Use of Data (CLOUD) Act streamlined the process by which foreign states can request data from U.S.-based firms. Through mutual legal assistance treaties, foreign governments are able to bypass the U.S. judiciary and directly request data from firms.[115] This legislation has facilitated increased cooperation between foreign governments and American firms, particularly in time-sensitive data requests, by removing administrative obstacles. Some critics, though, have argued that these public-private partnerships make it easier "for governments to escape the important safeguards that our legal systems have developed over time to protect political rights."[116]

Weaponized Interdependence

Following the Russian invasion of Ukraine in 2022, western states leveraged globalized trade, financial, and information networks to punish Russia for its aggression. Private technology firms, however, also participated in "weaponized interdependence" in the war in Ukraine via international information networks. Google, Meta, Microsoft, and Twitter all restricted the reach of Russian state-owned media outlets across their platforms, while PayPal and Apple suspended payment services within Russia. Additionally, Amazon Web Services halted new user sign-ups in both Russia and Belarus.[117] In Ukraine, Apple and Google played a role in protecting refugee evacuation routes by disabling live traffic features, and Airbnb waived booking fees for Ukrainian properties, facilitating direct financial support from international observers to property owners.[118] By occupying critical points in international

[113] Trugeur 2019.
[114] H. Farrell and Newman 2019.
[115] *Cloud Act Resources* 2019.
[116] Trugeur 2019.
[117] Tilley 2022.
[118] Airbnb 2022.

information networks, technology firms were able to shape the accessibility of information and services for participants in the conflict and for international observers.

2.6 The Utility of the Structural Perspective

While some scholars have downplayed the potential impact of DTD on international politics, asserting that it will mainly amplify existing domestic and global political trends rather than engender novel consequences, a structural lens challenges this perspective.[119] Through its unique attributes—encompassing engagement data, ease of duplication, and real-time high dimensionality—DTD empowers technology firms in a way that sets them apart from historical multinational corporations and hold consequences for broader dynamics within international politics.

Foremost, the framework proposed in this chapter pertains to theories concerning perception and misperception. Given their occupation of critical points in knowledge structures, technology firms shape the perception of events, policies, and political actors. The public's perception of these topics is particularly pertinent for the formation of audience costs. Media accessibility serves as a crucial factor in audience cost dynamics, as leaders can generate greater audience costs and thus make more credible commitments to foreign states when their own publics have access to robust media environments.[120] In the contemporary landscape, where populations increasingly rely on online sources for information, the mediation of information access by technology companies assumes a critical role in shaping public awareness of international events; broadly speaking, mass media can shape public attitudes and affect foreign policy decisions.[121] Consequently, this phenomenon bears implications for the generation of audience costs, the credibility of leaders' commitments, and, ultimately,

[119] Arsenault and Sarah E. Kreps 2022; Levy 2017.
[120] M. Baum and P. B. K. Potter 2015; Fearon 1995; Hyde and Saunders 2020.
[121] M. A. Baum and P. B. K. Potter 2019; M. A. Baum and P. B. Potter 2008; De Benedictis-Kessner et al. 2019.

the propensity for international conflict.

Moreover, this research adds to non-state actor theories in IR by highlighting a novel mechanism of non-state actor influence. While other types of non-state actors may seek influence through economic leverage,[122] lobbying,[123] norm creation and propagation,[124] or physical violence,[125] technology firms gain influence by harnessing insights from DTD. The adoption of this distinct mechanism by technology firms challenges conventional notions of state authority, as it enables them to undertake functions akin to those of states.

With this mechanism, technology firms exhibit dynamics most similar to those of transnational advocacy networks, in that they influence international outcomes by reshaping the information landscapes within which states operate.[126] However, a crucial distinction lies in their underlying motivations: whereas transnational advocacy networks strategically manipulate information environments to influence state actors and pursue social or normative objectives, technology companies alter these environments in their pursuit of profit maximization. Shoshana Zuboff refers to these dynamics as "surveillance capitalism," in which technology firms shape information environments and user behavior in ways which are economically lucrative.[127] Consequently, technology firms prioritize the dissemination of specific types of information over others—an example being the observation that Google search tends to favor content with American and mainstream attributes.[128]

In essence, technology firms impact international political dynamics as they navigate their path towards profit maximization. These insights emerge only when DTD is evaluated through a structural lens, revealing unique effects on international politics that are not mere accelerations of existing trends. These effects stem from the critical points in knowledge

[122]Grossman and Rossi-Hansberg 2006; Knorre, Kuchakov, and Skougarevskiy 2023.
[123]Alt et al. 1999; Grossman and Helpman 1992.
[124]Keck and Sikkink 1998.
[125]Weinstein 2007.
[126]Keck and Sikkink 1998.
[127]Zuboff 2019.
[128]Trielli and Diakopoulos 2022; Vaughan and Y. Zhang 2007.

structures that technology firms hold by virtue of their collection, storage, analysis, and communication of vast amounts of DTD. This occupation allows technology firms to shape information flows, influence public perceptions of political issues, and impact policy-making processes.

2.7 Conclusion

Despite its ubiquity in contemporary social and political life, DTD has received relatively limited attention in IR literature. The concentration of online activity within the services and devices of a few dominant firms grants them significant market control, allowing them to accumulate structural power. This power enables them to shape the informational environments of other actors and perform functions traditionally associated with states.

While DTD has gained prominence in political science literature as a means to conduct large observational studies, the substantive study of DTD lags behind. It is imperative to study DTD as it will only permeate further into social and political life. While digital data, and technology more generally, is not a sufficient condition to prescribe outcomes relevant to international politics, it often remains a necessary one. Existing theories regarding technology or the role of firms that emphasize instrumental or relational power fall short in addressing the structural power central to today's largest technology firms. Viewing these firms and the data they create, process, and transmit through a structural lens opens three main avenues for future research, respectively focusing on states, firms, and data.

The state-centric line of inquiry addresses the extent to which states use firms as assets in power competition. How might states use domestic technology firms to their advantage in such contexts? Following from Farrell and Newman (2019), the heterogeneous distribution of technology has consequences for state power relations.[129] While previous literature tends

[129]H. Farrell and Newman 2019.

to downplay the significance of firms in state power relations, technology firms, particularly those offering social networking platforms, provide a unique medium for the dissemination of ideology across borders through personal interaction. In the case of the U.S. and China, for example, scholars have argued that the United States aims to disseminate liberal ideology around the world, while China seeks to restrict its spread domestically.[130] When considering the control of information by today's technology firms, do these firms fundamentally differ from the MNCs of the past as an asset for power competition between states?

The second line of inquiry aims to explore the forms of power that firms accumulate through DTD and investigate the conditions that enable firms to leverage this power to coerce states. For example, Meta successfully exerted pressure on Australia, leading to a retreat from proposed content publishing regulations. However, similar attempts in Canada and France have been less impactful.[131] Furthermore, firms selectively cooperate with data sharing requests from state law enforcement and intelligence agencies. What are the conditions under which firms cooperate, and what are the consequences of selective cooperation?

Finally, further inquiry is needed into DTD itself. While scholars have conducted "audits" of online platform content, there is still a gap in the literature when it comes to systematically quantifying the heterogeneity of political content on the internet across different countries. What conditions are related to the similarity of political content between various geographic locations? Addressing this gap, the third chapter in this dissertation presents a comprehensive study that reveals significant variation in political content between countries, with a strong correlation to natural language. This finding challenges the implicit assumption in much of IR research that the internet provides a homogeneous information environment, with heterogeneity primarily driven by state censorship. However, it is well-known that content algorithms employed by large technology companies tailor information to individual users

[130] Mearsheimer 2018.
[131] Deschamps 2021; Rosemain 2021; Saba 2022.

based on their predicted interests, geographic location, and other factors. It is imperative for future research to address the implications of this heterogeneity.

DTD has demonstrated a consistent trend of expanding in its scope and scale. This trend is poised to persist as the demand for computational technologies continues to surge, driving an intensified push to generate and make use of more data. The continued integration of DTD into everyday life will significantly shape information environments central to international politics. Understanding how technology firms shape and mediate these information environments can provide insights into the evolving nature of state power relations in the digital age.

Chapter 3

All Internet is Local: Information Pools and Search Engine Representation of International Affairs

with Dr. Rochelle Terman

3.1 Introduction

Every day, billions of people use search engines to access information about the world. Search engines are powerful mediators in the public's exposure to information. According to the Pew Research Center, 86% of adults in the United States access news online at least some of the time (compared to 68% from television) and two thirds (65%) report using search engines like Google.[1] This trend transcends the United States. Platforms such as Google index vast amounts of digital content produced around the world, filter that content through opaque and proprietary algorithms, and deliver it to nearly every corner of the globe. Now more than ever, search engines, along with the algorithms they use to rank content, play a significantly influential role in the global transmission of information.

To the extent that international relations (IR) scholars have examined barriers to digital information transmission, they have focused largely on the actions of state actors that shape individuals' online experiences, such as digital censorship or propaganda production.[2] Surprisingly, search engines have received scant attention in IR research, despite their extensive reach and influence. Understanding how search engines represent international affairs is crucial, given that people rely heavily on platforms such as Google to gather information about both their own country and global politics. Moreover, Google tailors search engine results for the same query for different users, thereby presenting varying portrayals of topics and events to users based on their location, language, past browsing history, and other factors. In this manner, Google has the potential to contribute to cultural or political biases—there is a growing consensus that search engines like Google do not merely reflect reality, but play an important part in producing it. And yet, little research exists on how Google search results represent international affairs or how these representations are "localized" for users

[1] Shearer 2021.
[2] King, J. Pan, and Roberts 2013.

in different parts of the world.

This chapter conducts a large-scale audit of localized Google Search Engine Results Pages (SERPs) for queries related to international politics. Algorithm audits are a common technique used by researchers to evaluate bias and fairness in non-transparent and proprietary search algorithms, with past studies revealing discriminatory behavior towards both users and various types of content.[3] Employing a process that simulates user searches conducted in specific locales, we collect SERP data for a range of international affairs queries in 146 countries, generating hundreds of thousands of observations. Our technique allows us to identify distinct "information pools" that reflect particular configurations of readily available information accessible to users when they search for topics concerning world affairs.

We present three main findings. First, we find significant variation in the reach of ideological content, including state propaganda and material from transnational advocacy organizations. For example, searches regarding human rights may yield significantly different proportions of information from entities like Amnesty International or Chinese state media, depending on the user's location and language. Second, we find that localized results strongly correlate with search language, suggesting language as a primary mediator of people's exposure to information about international affairs. Finally, we trace how SERPs change in response to salient events. Analyzing results related to the war in Ukraine generated both before and after the 2022 Russian invasion, we find that geographic clustering in the content of SERPs becomes more substantial following the invasion, particularly among lesser-spoken language groups.

These findings have significant implications for the field of IR. For many scholars, internet technologies promised to make the world more open, connected, and democratic by enabling the flow of information across national borders. In addition to state actors engaging in practices such as digital censorship and propaganda production, we identify another barrier

[3]Bandy 2021.

to information that resides outside of direct state control and is facilitated mainly by private actors: search engines. By exposing the biases and variations produced by personalization, we shed light on the ways in which search engines shape representations of international affairs. Although we do not directly address the effects of SERPs, this study identifies a key mechanism whereby people systematically encounter different information regarding world politics based on their search language and country of origin.

3.2 Search Engines as Social Power

Because the amount of information on the web is much larger than any person can navigate, search engines like Google act as essential "information mediators" between society and the Internet.[4] By filtering, ranking, and directing users towards "relevant" content, these platforms hold tremendous influence over our exposure to information. Far from acting as neutral aggregators, search engines operate more as gatekeepers and curators, wielding the ability to "grant visibility and certify meanings" about the social world.[5]

We examine the most powerful digital information mediator operating today: Google. Google.com is the most visited site on the Internet (followed by Youtube, another Google product), amassing over 89 billion monthly visits in 2022.[6] Indeed, it is no accident that "googling" is synonymous with searching the web, as the firm captures over 90 percent of the search engine market share. Other search engines have emerged in particular locales—such as Baidu in China and Yandex in Russia—but have limited reach beyond national borders.[7] In contrast, Google dominates the global search market, capturing the majority of searches in most countries.

As search engines have become ubiquitous in our daily lives, both scholars and poli-

[4]Bandy 2021.
[5]Gillespie 2017; Diakopoulos et al. 2018.
[6]Mohsin 2022.
[7]For a comparative study of Baidu and Google, see (M. Jiang 2014).

cymakers have become increasingly concerned about their potential to present biased or discriminatory information. Motivated by this concern, a burgeoning literature seeks to understand the operations of search engines and their societal effects.[8] Due to the opaque and proprietary nature of search algorithms, researchers typically use *algorithm audits* to scrutinize search engines and uncover potential biases.[9] These audits have addressed a variety of components within Google search, including the snippets that summarize webpages,[10] news and top stories,[11] knowledge graphs,[12] maps,[13] images,[14] autocomplete,[15] and search suggestions.[16]

Together, these studies demonstrate that search engines in general—and Google in particular—systematically privilege certain kinds of content over others.[17] Audits find that search engines tend to prioritize digital content that is more American, more commercial, and more popular.[18] One study finds that the top 20% of domains accounted for 96% of all search results in queries related to politics.[19] A similar trend emerges within Google News and its "top stories" section, where a handful of sources account for the vast majority of impressions.[20]

As a result, search engines have the potential to perpetuate existing hierarchies in knowledge production. For example, one study examined localized versions of Google to find that some locations—particularly those with highly developed publishing and scientific industries—feature more locally-produced results, while others feature content produced else-

[8]Granka 2010; Introna and Nissenbaum 2000; Laidlaw 2009; Noble 2018.
[9]Sandvig et al. 2014; Bandy 2021.
[10]Hu et al. 2019.
[11]Trielli and Diakopoulos 2019; Fischer, Jaidka, and Lelkes 2020; Robertson, Jon Green, et al. 2022; Robertson, Lazer, and Wilson 2018.
[12]Robertson, Lazer, and Wilson 2018.
[13]Kliman-Silver et al. 2015.
[14]Kay, Matuszek, and Munson 2015; Noble 2013.
[15]Robertson, Lazer, and Wilson 2018.
[16]Ibid.
[17]Introna and Nissenbaum 2000.
[18]Lawrence and Giles 1999; Mowshowitz and Kawaguchi 2005; Vaughan and Thelwall 2004; Vaughan and Y. Zhang 2007.
[19]Robertson, Lazer, and Wilson 2018.
[20]Trielli and Diakopoulos 2022; Fischer, Jaidka, and Lelkes 2020.

where.[21] While the precise causes for such biases remain unknown, they may arise from the metrics used by search engineers, who often prioritize "customer satisfaction" and "relevance" over fairness and representation.[22]

The potential for bias within search results is particularly concerning in the context of political queries, such as information on electoral candidates or polarizing topics.[23] One study finds that Google curates information about American political candidates through various editorial choices.[24] Another discovers that Google Search's snippets tend to amplify partisanship relative to the pages they intend to summarize.[25] Robertson and colleagues collected SERPs and autocomplete suggestions from participants searching for politically-related information a month after Donald Trump's inauguration, finding substantial variation in both content and personalization across query type, user characteristics, and time.[26] Using a similar approach, a different study on political searches found that results ranked lower on Google are more left-leaning, on average, than those positioned higher.[27] Most recently, Trielli and Diakopoulos (2022) discovered that Google search returns a common set of results to disparate political queries, exerting a "mainstreaming effect" by neutralizing partisan information-seeking.[28]

These findings are important by virtue of the power of rankings to shape human attention. Users are more likely to click on a result, and believe it is more relevant, the higher it is ranked.[29] In one survey, 90% of respondents said they were likely to click on the first page of

[21] Ballatore, Graham, and Sen 2017.
[22] Van Couvering 2007.
[23] Metaxa et al. 2019; Trielli and Diakopoulos 2022; Hu et al. 2019; Robertson, S. Jiang, et al. 2018; Robertson, Jon Green, et al. 2022; Metaxas and Pruksachatkun 2017; Diakopoulos et al. 2018; Kulshrestha et al. 2017.
[24] Diakopoulos et al. 2018.
[25] Hu et al. 2019.
[26] Robertson, Lazer, and Wilson 2018.
[27] Robertson, S. Jiang, et al. 2018.
[28] Trielli and Diakopoulos 2022.
[29] B. Pan et al. 2007.

results.[30] In observational research, the top three search results consistently receive over 50% of clicks, with 75% of clicks being made on the first page.[31] More generally, search engines are widely seen as neutral purveyors of information; most users trust the quality of search results and remain unaware that the results are curated or personalized.[32]

As expected, several experimental studies show that search engines significantly shape users' attitudes, beliefs, and behavior—a dynamic Epstein and Robertson (2015) coin as the "search engine manipulation effect" (SEME).[33] Using mock search results surrounding the 2014 Indian elections, they find that nearly a quarter of undecided voters could be swayed by biased search results, even when exposed to interventions to suppress SEME.[34] Similar effects are observed for search results related to science, health, as well as race and gender stereotypes.[35]

Complicating this picture is the fact that Google search results have become increasingly personalized, aimed at providing more relevant content to users. The ranking algorithm incorporates user-provided information, such as location or past browsing history, to modify the set or order of results. Consequently, individuals searching with identical queries often receive different results.[36] For example, one study found that some users searching "Tahrir Square" received news covering the Egyptian Revolution, while others saw links to travel agencies.[37] Rising personalization raises concerns about the potential for "filter bubbles" or "echo chambers," wherein users are systematically exposed to biased information that reinforces their preexisting political and social attitudes. While the "filter bubble" hypothesis remains controversial, one study estimates that around 12% of Google search results in the

[30] Mohsin 2022.
[31] *Google Organic CTR History* 2022.
[32] B. Pan et al. 2007; Tripodi 2018; Epstein and Robertson 2015.
[33] Epstein and Robertson 2015.
[34] Epstein and Robertson 2015; Epstein, Robertson, et al. 2017.
[35] Knobloch-Westerwick et al. 2015; Mejova, Gracyk, and Robertson 2022; Allam, Schulz, and Nakamoto 2014; Kay, Matuszek, and Munson 2015; Noble 2013.
[36] Xing et al. 2014; Hannák et al. 2017; Kliman-Silver et al. 2015; Ochigame and Ye 2021.
[37] Pariser 2011.

United States differ due to personalization.[38] Personalization also operates along language lines, whereby search algorithms expose users to distinct resources that may contain specific biases or stereotypes unique to their search language.[39]

Thus far, however, most of the research on personalization has been limited to U.S. contexts, overlooking how it unfolds around the world.[40] Given their global reach, knowing how search engines distribute information worldwide is of crucial importance, especially for understanding the relationships between informational environments, public opinion, and international politics. Surprisingly, to our knowledge, no prior research has examined the representation of international affairs in search engine results, or how such representations vary across countries, regions, or languages.

3.3 Localization and Informational Pools

While research on search engine personalization is still in its early stages, the strongest and most consistent finding concerns the role of location. More than any other feature, users' geolocation—inferred by IP address and other factors—exerts the greatest influence on differences in search results.[41] For simplicity, we refer to location-based personalization—i.e., the variation in search results by location—as *localization*. The motivating principle behind localization is straightforward. When users search for "cafe" or "movie theaters," they are likely interested in finding locations nearby. Indeed, users living mere blocks from one another will likely find different results for the same term, especially when searching for physical places and local establishments.[42]

[38]Kliman-Silver et al. 2015; Robertson, S. Jiang, et al. 2018; Robertson, Lazer, and Wilson 2018; Robertson, Jon Green, et al. 2022.

[39]Luo, Puett, and Smith 2023.

[40]Hannák et al. 2017; Kliman-Silver et al. 2015; Robertson, Lazer, and Wilson 2018; Robertson, Jon Green, et al. 2022; Hu et al. 2019.

[41]Xing et al. 2014; Kliman-Silver et al. 2015.

[42](Kliman-Silver et al. 2015) One U.S.-based study finds that queries for political and controversial topics tend to be more consistent across space, while queries for local terms ("airport") are highly personalized and

User-level personalization of search results—which occurs as a function of localization and other factors—produces what we call *information pools*. Information pools are distinct configurations of informational resources that users readily encounter when they seek information on a particular topic. Unlike echo chambers or filter bubbles—which tend to be conceptualized as reflective of a particular ideological position or viewpoint (e.g., liberal or conservative)—information pools are broader phenomena that encompass *all* informational resources available to users, including websites, news articles, autocomplete suggestions, and images. Localization plays a pivotal role in shaping these information pools, as it directly contributes to the creation of personalized search engine outcomes.

Google uses several data sources to determine searchers' locations, some of which are only accessible if users consent to share them. For example, on mobile devices, Google can use GPS to pinpoint a device's location with remarkable accuracy.[43] If users are logged in to their Google account, search results may consider their location history, browsing history, and "labeled places" in Google Maps. While the sharing of location data remains a contentious issue, users now possess the ability to manage many information sources through various permission settings.

However, one source of location data generally remains beyond the control of users: IP address. IP addresses establish connections between devices to websites and are a necessity for internet usage. An Internet Service Provider assigns an IP address to a device, generally corresponding to its geographic location. Consequently, Google can estimate a user's location based on their IP address, irrespective of their personal settings. As Google itself puts it, "You can't prevent apps or websites that you visit, including Google, from getting the IP address of your device, because the Internet does not work without it. This means all apps

differ widely.

[43](Ghose, Goldfarb, and Han 2013) Not surprisingly, effects of localization may be more pronounced on mobile search results.

and websites that you visit can usually infer some information about your location."[44]

Indeed, Google openly acknowledges that it *always* estimates the general area from where users are searching (defined as an area larger than 1 square mile with at least 1000 users). Using this information, Google search automatically directs users to a localized version of its search engine (e.g., google.fr), sets a default language (e.g., French), and returns results that are personalized for users in that general location.[45]

To those outside of the firm, Google's search algorithm remains an opaque process, and little is known about the ways it selects and ranks information for users. Nonetheless, we identify three sets of factors that interact to produce localization. The first pertains to the users themselves. Search engines like Google strive to provide results that are as "relevant" as possible. To do so, they consider numerous factors related to the end user and her context, including her primary language, her interests in particular topics or sources, and her past searches. Even if Google knows little about this *particular* user, the search engine aggregates huge amounts of data from other, similar users to deduce the relevance of particular results. Insofar as this input data varies across geographic locations (since user behavior also varies by location), search engines will produce different results in different locales.

The second set of factors pertain to the content that is indexed, filtered, and ranked by search engines. Search engines like Google introduce significant market pressures for online publishers due to their role in directing internet traffic. According to the Parse.ly external referrer dashboard in April 2022, Google Search accounted for 56.6% of all external traffic to online publishers. Given search engines' centrality within the realms of media and advertising, search engine optimization (SEO) has emerged into a multibillion-dollar industry. The primary aim of SEO is to enhance a website's visibility and prominence in search outcomes, spanning a spectrum from general searches to those of specific demographics, such as users

[44] *How Google Protects Your Privacy & Keeps You in Control* 2022.

[45] *Understand & Manage Your Location When You Search on Google - Computer - Google Search Help* 2022.

in a particular geographic location. This endeavor drives content publishers to adapt their material to align with a search engine algorithm's criteria for prioritization, effectively creating a "supply side" to content production. In this way, search engines exert tremendous influence over the very content they are designed to organize, resulting in systematic shifts in digital content creation.[46] Beyond publishers, third parties often seek to take advantage of search algorithms for their own (often political) purposes, thereby prompting search engines to counteract such behavior.[47] Essentially, a user in a specific location may encounter a distinct set of search results due to publishers (or third party spammers) using certain tactics to target particular user groups.

The third set of factors resides in the algorithm itself, encompassing the procedures and decisions it uses to deliver "relevant" results. Google's ranking algorithm is not a neutral aggregator of information; it hinges on particular choices that systematically prioritize certain kinds of content based on predictions of content relevance. Search engine algorithms are constantly evolving as engineers tune their optimization strategies, fend off spammers and malicious actors, and adapt to technological innovation. Due to the proprietary and trade-secret nature of search engines, the algorithms they use to filter and rank content are essentially unknowable to outsiders.

Because exposure to online content impacts individual political views, the primary significance of this research for the field of international relations hinges on the role of the public in influencing state behavior on the international stage.[48] The influence of the public in international politics has been well studied. Leaders, for instance, make concessions on foreign policy matters to align with public sentiment, with effects seen in military force us-

[46] Vaughan and Thelwall 2004.
[47] Metaxas and Pruksachatkun 2017; Metaxas and Mustafaraj 2009.
[48] Bryson 2020; Y. Kim, H.-T. Chen, and Gil De Zúñiga 2013; Warner 2010.

age,[49] trade policy,[50] and climate policy adoption.[51] Localization assumes an important role in international politics due to its potential to differentially shape public opinion between different locations. Baum and Potter (2019) contend that this phenomenon grants more power to political leaders.[52] The fragmented information landscape in contemporary media reduces the likelihood of the public encountering information that contradicts their beliefs, thereby solidifying their support for leaders and disapproval of opposition.

Moreover, the impact of localization is especially amplified in the context of extremist content, as the efficacy of online platforms' content moderation systems varies across locations and languages. While not strictly a search engine, it is worth noting that documents disclosed in 2021 indicate that 87% of Meta's global content moderation budget was allocated to the United States, and content moderation was available in only around 70 of the platform's over 100 supported languages.[53] Deficiencies in content moderation systems have been associated with the spread of inflammatory information during civil conflicts,[54] and extremist online content has been linked to radicalization and mobilization in support of extremist causes.[55]

The sheer complexity of search engines—to say nothing of their inscrutability— fundamentally limits our ability to make causal claims. As a sociotechnical system, these platforms involve opaque algorithms interacting with user-level and content-level inputs in a constant feedback loop.[56] Identifying the causes of SERP bias would require teasing apart the influence of users, content publishers, and the ranking algorithm itself, a challenging if not impossible task.[57] Nonetheless, our descriptive findings provide an initial first step towards

[49]Tomz, Weeks, and Yarhi-Milo 2020.
[50]Kono 2008.
[51]Bromley-Trujillo and Poe 2020.
[52]M. A. Baum and P. B. K. Potter 2019.
[53]*Form 10-K* 2022; Turowski 2023.
[54]*Human Rights Impact Assessment* 2018; Olson 2022.
[55]J. Carter, Maher, and Neumann 2014; Mitts 2019; Vidino and Seamus 2015.
[56]Gillespie 2017.
[57]Kulshrestha et al. 2017; Ørmen 2016.

understanding how search engines shape the information pools representing international politics in different parts of the world.

3.4 Data and Methods

To examine the representation of international affairs in Google search results, we conduct a series of Google searches across different geographic locations and languages. To accomplish this, we use SerpAPI, an enterprise API specifically designed for SEO applications. SerpAPI employs proxy servers and leverages Google's geolocation parameters to gather search results from various locations around the world.[58] Effectively, SerpAPI enables users to simulate searches originating from a preferred location, thereby disguising their actual location.

3.4.1 Search Workflow

Our study involves conducting Google searches from a total of 146 countries. The selection of these countries is based on two criteria: inclusion in Google Ads for geographic targeting and support for the most spoken language in that country through Google Translate. These criteria are essential for employing geographic targeting within SerpAPI and facilitating the translation of the search queries from English to the local language. SerpAPI allows for localization of queries to sub-national locations supported by Google Ads, such as regions, cities, or even specific neighborhoods within urban areas. To determine the search location within each country, we use the following methodology:

1. when available, we use the primary Google Ads location for the capital city;
2. in cases where the primary Google Ads location for the capital city is unavailable, we use the primary Google Ads location for the region, state, or province in which the capital city is located;

[58] *Google Search API* 2022.

3. if neither the primary Google Ads location for the capital city nor the regional location is available, we use to the primary Google Ads location for the country as a whole.

In all cases, we conduct searches in the most widely spoken language of each country. We opt to use the most spoken language instead of the country's official language to ensure that our approach best approximates how typical users would search for information. We begin with a list of queries in English, which we then translate into the 70 local languages within our study's scope using Google Translate. Furthermore, we modify several Google interface parameters to align with local contexts. Specifically, we search on the country-specific Google domain, such as www.google.com.mx for Mexico, and configure the interface language to match the search language—a criterion that Google employs in its search result ranking.[59] When a country lacks a unique Google domain or Google does not provide interface support in the search language, we revert to using www.google.com and the English interface. Other parameters are left to their default settings. For a table of all search locations, languages, and interface parameters, please refer to Appendix A.1.

3.4.2 Auditing Strategies

There are a variety of ways in which an individual might use Google to seek out information on international affairs. To capture different user behaviors, we develop two distinct auditing strategies. The first strategy focuses on users seeking information about salient topics or events related to international affairs, such as a social issue or an international conflict. To emulate this user behavior, we conduct searches on subjects such as human rights issues, COVID-19, and the war in Ukraine. Using Google Search and Google Images, we execute separate searches from every country for each query listed in Table 1. We conducted two rounds of Google searches on February 5 and March 9, 2022, along with Google Image

[59] *How Google Determines the Language of Search Results* n.d.

searches on February 6 and March 10. This process involved repeating the complete set of queries during each round. In each case, we collected the first page of organic search results from Google Search and the first five image results from Google Images.

Table 3.1: Topic and Event Search Queries

Queries	
freedom of assembly	Ukraine crisis
freedom of religion	covid lab-leak theory
freedom of the press	coronavirus lab-leak theory
freedom of movement	Uyghurs genocide
freedom of expression	Uyghurs terrorism
Russia invasion	

The second auditing strategy revolves around users seeking insights into the culture, politics, or society of foreign countries. Termed "worldview" searches, we search the names of 193 countries from each country in our universe.[60] Once again, we conducted two rounds of searches in February and March 2022, collecting data from the first page of organic Google Search results and the first five Google Image results.[61] Table 2 summarizes both auditing strategies.

3.4.3 Measuring Informational Pools

With this data, we assess the presence of information pools using two primary strategies. We first examine variation in the sources returned by Google search. SERPs deliver content

[60]Our search coverage is wider in terms of the number of countries we can search about compared to the number of countries we can search from. This discrepancy arises from constraints imposed by SerpAPI and Google Translate when performing geolocated searches. We encountered 47 countries that were either not supported by Google Ads or whose most spoken language was not supported by Google Translate. However, by incorporating these 47 countries along with the initial 146 countries in our search universe, we expanded our search scope to encompass a total of 193 countries.

[61]For Google Search, we completed the first round from February 16-18 and the second round from March 16-17. For Google Images, we completed the first round from February 18-21 and second round from March 17-19. To address the significant length of time needed to conduct a full round of searches, we completed all searches in order of search query, not search origin. By doing so, we could complete all searches for a given query in under 10 minutes before proceeding to following the query.

Table 3.2: Summary of Auditing Strategies

Auditing Strategy	Round 1 Dates	Round 2 Dates	Total Locations	Total Queries	Total Searches	Total Observations
Salient Topics and Events	Feb 5-6, 2022	Mar 9-10, 2022	146	11	6,412	43,238
Worldview	Feb 16-21, 2022	Mar 16-19, 2022	146	193	108,764	950,434

produced by a wide array of publishers, including established international human rights organizations (e.g., Amnesty International), international organizations (e.g., United Nations), governments, media outlets, and private companies. Variation in the reach of these publishers is important because different kinds of actors produce different messages about political topics. For instance, regarding human rights topics, resources from Amnesty International or UNHCR are likely to convey substantially different information compared to materials generated by an authoritarian government.

To identify publishers, we trim the URL of each result to its main domain, removing any language or location-specific prefixes or suffixes.[62] We then evaluate the similarity between SERPs by calculating the proportion of shared publishers. In Appendices A.2 and A.3, we provide supplementary analyses introducing an additional metric for SERP comparison: rank biased overlap (RBO).[63] RBO is a measure of ranked listed similarity and upweights the importance of top list entries, aligning with the behavior of users who primarily interact with top-ranked results.[64]

Differences among publishers within SERPs hold significance when the content produced by those publishers varies in meaningful ways. To measure variation in SERP content, we leverage the results obtained from Google Images. Examining image variation permits direct comparisons of SERP content, regardless of the specific publishers or language of a

[62]For example, www.bbc.com and www.bbc.co.uk are both trimmed to 'bbc', and www.cnn.com and www.cnnespanol.cnn.com are both trimmed to 'cnn'.

[63]Webber, Moffat, and Zobel 2010.

[64]RBO has been used in other Google search audits, such as (Robertson, Lazer, and Wilson 2018).

SERP. We assume that different image content between two SERPs implies distinctiveness in the remaining information within those SERPs. This assumption allows us to leverage image-based comparisons as an indicator of overall information variation within the SERPs.

To quantify image content, we employ image embeddings generated by a Vision Transformer (ViT) model, pretrained on the ImageNet-21k database.[65] ViT is a computer vision model that uses a transformer architecture, which has become common in both computer vision and natural language processing applications. We use ViT to generate a 768-dimensional embedding for each image obtained from the Google Image results. The pretraining of ViT on the ImageNet-21k database, comprising 14 million images categorized into 21,843 classes, allows the model to encode distinguishing features into the embeddings that are useful in discerning the various classes in the data set. This enables us to measure image content and characteristics.

We introduce a metric, Maximum Cosine Similarity (MCS), to evaluate the similarity of the image embeddings between two SERPs. MCS starts by computing the cosine similarity between every pair of image embeddings from the two SERPs. It then iteratively identifies the most similar pair of images, records their cosine similarity, and excludes them from further consideration, repeating this process until five pairs are selected. MCS sums these five cosine similarity values to indicate the overall similarity of the two SERPs. In Appendices A.2 and A.3, we present supplemental analyses with a second metric second metric called Cosine Similarity Rank Biased Overlap (CS-RBO), which considers ranking among the top five results. Further details regarding the MCS and CS-RBO metrics are presented in Appendix A.4.

[65] Dosovitskiy et al. 2021.

3.5 Findings

3.5.1 Publisher Disparities in Localized SERPs

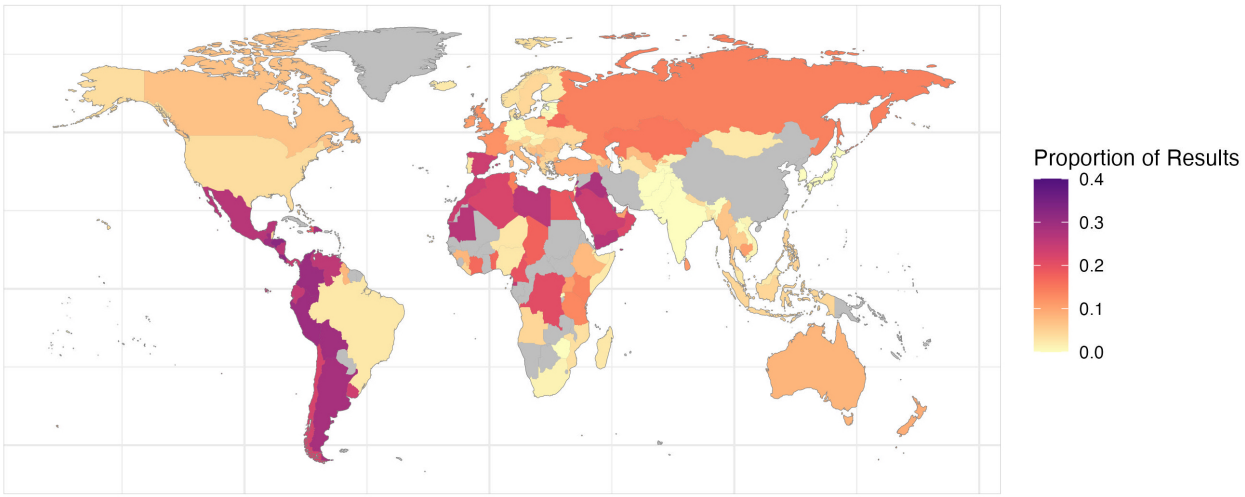
We observe significant variation in the presence of different publisher types in localized search results. We first consider searches for topics related to human rights, creating a concise list of prominent transnational advocacy organizations that publish content concerning human rights issues.[66] Figure 3.1 displays the proportion of results from these publishers across five queries: "freedom of expression," "freedom of movement," "freedom of the press," "freedom of assembly," and "freedom of religion." Our findings indicate that results from transnational advocacy organizations are predominantly concentrated in the Americas and the Middle East and North Africa, particularly in countries where Spanish and Arabic are commonly spoken. Notably, in populous countries such as India, Pakistan, and Japan, these publishers are entirely absent from the search results. This means that users conducting searches from these countries are unlikely to encounter any information from these specific publishers on the first page of search results. This language-aligned variation persists even among geographic neighbors. For example, in Brazil, the top three "freedom of expression" search results originate from FIA, InfoEscola, and Aurum—respectively, a Brazilian business school, a free education website, and a legal software company. Conversely, in neighboring Argentina, the top three publishers for the same query are Amnesty International, Wikipedia, and the Organization of American States Inter-American Commission on Human Rights.[67]

Next, we explore the presence of Xinhua, the official state news agency of the People's Republic of China, in SERPs related to COVID-19. Throughout the COVID-19 pandemic, Xinhua promoted a pro-government narrative and downplayed the social and economic con-

[66]This list includes: Amnesty International, the Office of the United Nations High Commissioner for Human Rights, Human Rights Watch, UNESCO, Freedom House, Frontline Defenders, Equality and Human Rights Commission, and the ACLU.

[67]These results reflect the first wave of searches in February 2022.

Figure 3.1: Prevalence of Transnational Advocacy Organizations in Freedom Searches

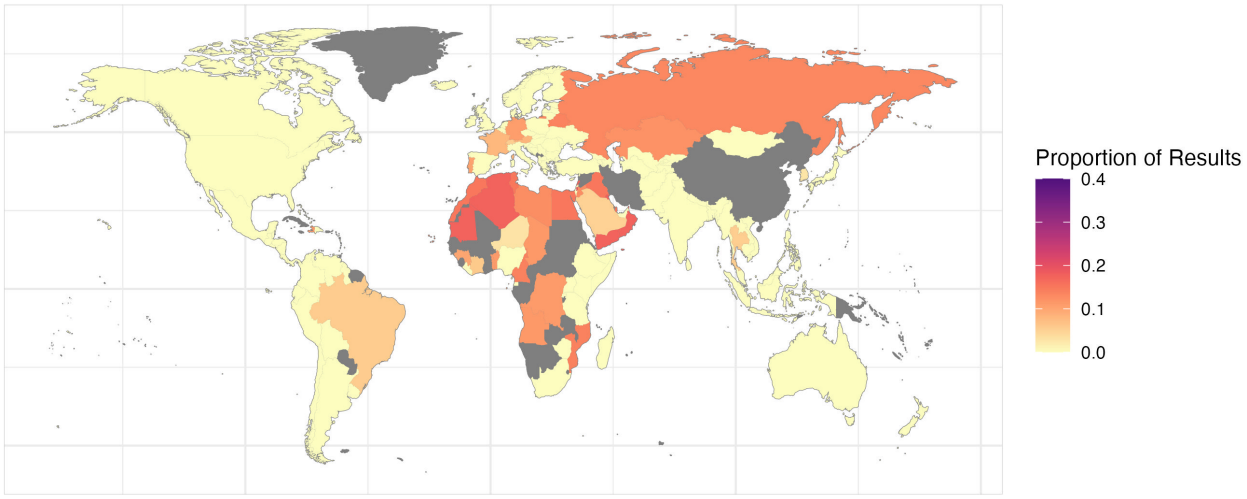


sequences of the disease.[68] Figure 3.2 illustrates the proportion of search results originating from Xinhua for the queries "covid lab-leak theory" and "coronavirus lab-leak theory." Results from Xinhua are most concentrated in Africa and the Middle East, aligning with previous research highlighting China's substantial endeavors to cultivate favorable media coverage in the region.[69] This finding suggests that people in this region are more prone to encounter information sympathetic to China.

It is important to emphasize that the causal mechanisms underpinning the variation in the reach of particular publishers remain opaque. However, one plausible explanation relates to the supply side of content production. Publishers often generate content in specific languages tailored to their target demographics. For example, Amnesty International produces news in English, Spanish, French, and Arabic, whereas OHCHR supports the same languages as well as Simplified Chinese and Russian (Figure 3.3). Xinhua, as a state news agency, also publishes materials in multiple languages, possibly to target specific audiences worldwide. It is clear from our analysis that the representation of publishers promoting

[68] El Damanhoury and Garud-Paktar 2021; Z. Chen and Xu 2021.
[69] Eisenman 2023; Xin 2009.

Figure 3.2: Covid Lab-Leak Theory Search Results from Xinhua



ideology and/or propaganda regarding international issues is far from uniform across different locales, substantially influencing the accessibility of information for users in certain countries and regions. While some governments may censor certain ideological content on their national internet, we suspect the variation depicted above is largely driven by search algorithms operating independently of state control.

Figure 3.3: Language-Specific Content Production of Amnesty International and OHCHR

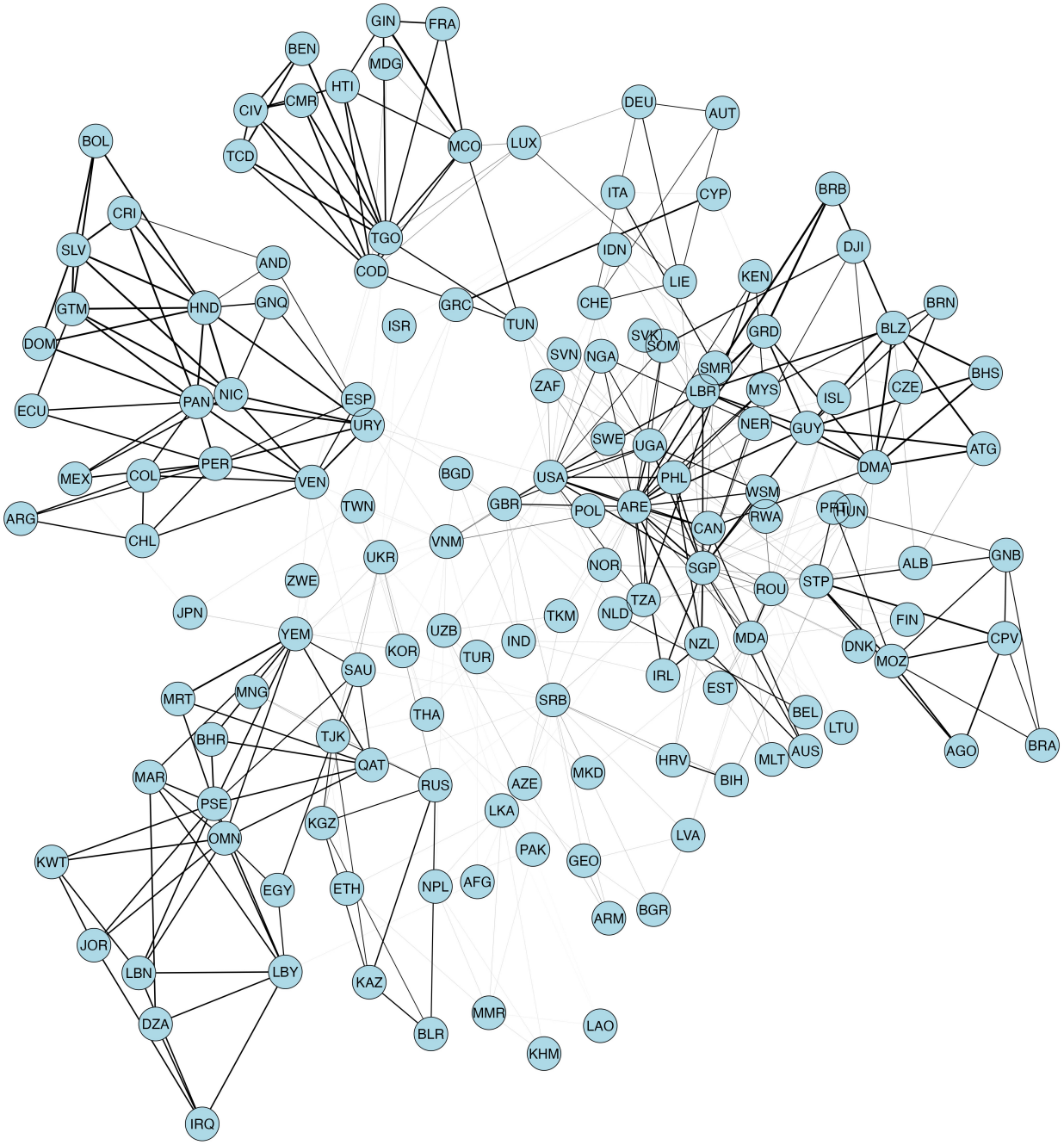
3.5.2 Language, Information Pools, and SERP Clustering

Which countries encounter similar content in their SERPs? We leverage the results from the worldview auditing strategy to construct a network of countries, visualizing their proximity based upon SERP similarity (Figure 3.4). In the network, each node represents a country from which we conducted localized Google searches. The edges are weighted by the mean proportion of shared sources across all queries and their respective SERPs for the connected countries. We depict an edge only between each node and its three most similar nodes in the network, determined by the highest mean proportion of shared sources.

In the network, notable groupings emerge for Spanish and French-speaking countries in the upper left section, along with a somewhat discernible cluster of Arabic-speaking countries in the lower left. These clusters suggest shared source similarity within each group, as well as disparities in sources between these groups and countries outside of the cluster. In other words, sources that appear within SERPs from countries in the Spanish, Arabic, and French clusters are generally absent from other language groups. In contrast, the English-speaking cluster in the upper right appears less defined, encompassing several countries representing smaller language groups, such as Kenya (Swahili) and Poland (Polish). This suggests that to the extent that countries within these smaller language groups receive information from sources commonly associated with other languages, they receive the most information from Anglophone publishers. This pattern is consistent with other research indicating that search biases reproduce established knowledge hierarchies.[70] Additionally, countries with less commonly spoken languages, including Serbia (Serbian), Turkey (Turkish), and Israel (Hebrew) appear isolated in the network, indicating that their SERP sources are not prevalent in other countries. Appendix 3.2 provides reconstructions of this network based on the RBO metric as well as the Google Image results.

[70] Ballatore, Graham, and Sen 2017.

Figure 3.4: Network of Shared Sources



3.5.3 Correlates of SERP Variation across Countries

What factors explain SERP similarity between countries? We continue to investigate the results of our worldview searches by presenting two regression models below. It is important

to re-emphasize that our analysis does not make causal claims, as determining the precise causal mechanisms producing search engine results is a complex or perhaps impossible task. Instead, our aim is to provide a descriptive analysis to highlight the correlates of SERP variation.

The following models contain different dependent variable specifications to evaluate the similarity between two SERPs. In Model 1, the dependent variable is the mean proportion of shared sources across all SERPs for the dyad. Model 2 leverages the Google Image results, with the dependent variable being the mean MCS value across all queries for the dyad.

We include six independent variables in the models. The first set of variables focuses on the supply side of content production, whereby publishers produce content in particular languages for certain target audiences. The first variable, *Common Language*, is a binary indicator for whether the two countries share an official language. The second variable, *Total Bilateral Trade Flows (log)*, represents the natural log of the sum of trade flows between the two states. We hypothesize that countries sharing a common language or displaying stronger economic ties are more likely to be targeted by the same content publishers.

Next, we consider social and economic variables that address different algorithmic aspects of SERP variation. *Colonial History* is a binary indicator that signals whether the two states share any colonial ties. States with a shared colonial history often have cultural similarities, which can influence the content present in SERPs. *Social Connectedness* represents the dyad's value obtained from the Facebook Social Connectedness Index (SCI) data set, which measures the proportion of all Facebook users between the two countries who share a Facebook friendship.[71] Facebook friendships have been shown to correspond closely to real-world friendship networks.[72] Given the tendency for homophily in social networks, we posit that SCI can serve as a reasonable proxy for the similarity of user-level characteristics

[71] M. Bailey, Cao, et al. 2018.
[72] Yearwood et al. 2015.

and interests across countries.[73] To align with previous literature, we take the natural log of the SCI variable.[74]

Additionally, we incorporate *GDP Difference (log)*, which takes the natural log of the difference in GDP of the two states. This variable accounts for potential impact of websites tailoring content based on users' device types, influencing their rankings on search engines. By including GDP differences, we can approximate the connection between device type and SERP content. Finally, *Same Region* is a binary indicator that addresses localization, specifically indicating whether two countries belong to the same United Nations sub-region. Table 3.3 presents the results of these models.

[73]McPherson, Smith-Lovin, and Cook 2001.
[74]M. Bailey, Gupta, et al. 2021; M. Bailey, P. Farrell, et al. 2020.

Table 3.3: Correlates of SERP Variation

	Dependent Variable	
	Shared Sources (Pct)	Image Max CS
	(1)	(2)
Common Language	0.255***	0.145***
	(0.005)	(0.004)
Total Bilateral Trade Flows (log)	−0.001	0.004***
	(0.001)	(0.001)
Colonial History	0.059***	0.027***
	(0.012)	(0.010)
Social Connectedness	0.007***	0.011***
	(0.001)	(0.001)
Same Region	0.048***	0.028***
	(0.005)	(0.004)
GDP Difference (log)	−0.001	0.002**
	(0.001)	(0.001)
Constant	0.164***	0.279***
	(0.023)	(0.021)
Observations	7,022	7,430
R^2	0.429	0.278
Note:		*p<0.1; **p<0.05; ***p<0.01

The effect of *Common Language* is positive and statistically significant at conventional levels in both models, indicating that countries sharing the same language tend to have more similar SERPs, both in terms of publishers and image content. In Model 1, the coefficient of 0.255 suggests that SERPs from countries with a shared language have approximately 25% more shared publishers, equivalent to about two additional results per page on Google Search. Similarly, the positive coefficient of *Colonial History* in both models indicates that countries

with shared colonial ties exhibit greater similarity in their SERPs. *Total Bilateral Trade Flows* only reaches conventional levels of statistical significance in Model 2, suggesting that trade relationships are associated with differences in SERP content but not in the publishers comprising the SERPs.

The positive coefficient on *Social Connectedness* in both models indicates that countries with stronger social ties tend to have more similar SERPs. We take this as evidence of user-level personalization of search results. The coefficient in both models is also positive for *Same Region*, which meets our theoretical expectation that localization contributes to similarity of SERPs. Finally, the coefficient on *GDP Difference* only reaches conventional levels of statistical significance in Model 2, albeit with a relatively small effect size. Given the inconsistent results between the models and the further inconsistency for this variable in alternative specifications (presented in Appendix A.3), we approach the evidence cautiously and hesitate to conclude that GDP similarity, and by extension device type, has a positive association with similarity in content publishers.

Overall, these results indicate that language is the strongest correlate of SERP similarity between countries, exhibiting a strong relationship with both dependent variables. While location and other factors also exhibit associations with SERP similarity, their predictive strength appears comparatively weaker to that of language. Additional models with different dependent variable specifications, as detailed in Appendix A.3, further support the robustness of these results.

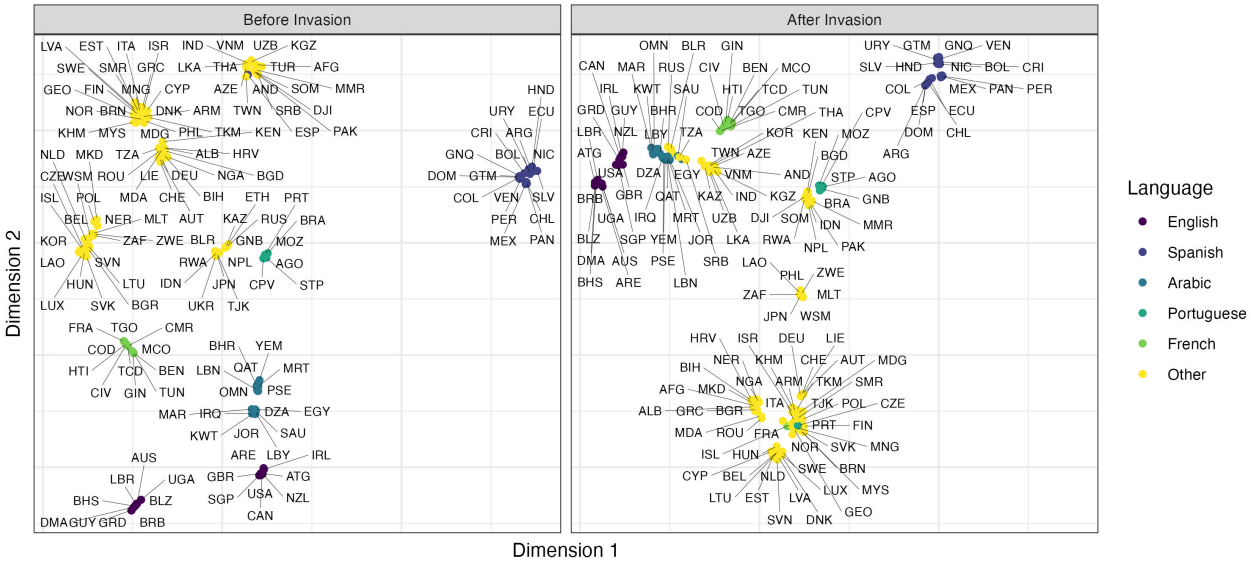
3.5.4 How SERPs Change with International Events: Russia's Invasion of Ukraine

How do SERPs change over time? A significant event that occurred during our data collection period was the Russian invasion of Ukraine on February 24, 2022, falling between

our two rounds of data collection in February and March. Given that our auditing strategy related to salient topics and events specifically included the terms "Ukraine crisis" and "Russia invasion," we can investigate changes in SERP content following the invasion. Our analysis centers on investigating whether countries formed clusters based on their SERP content both before and after the Russian invasion of Ukraine.

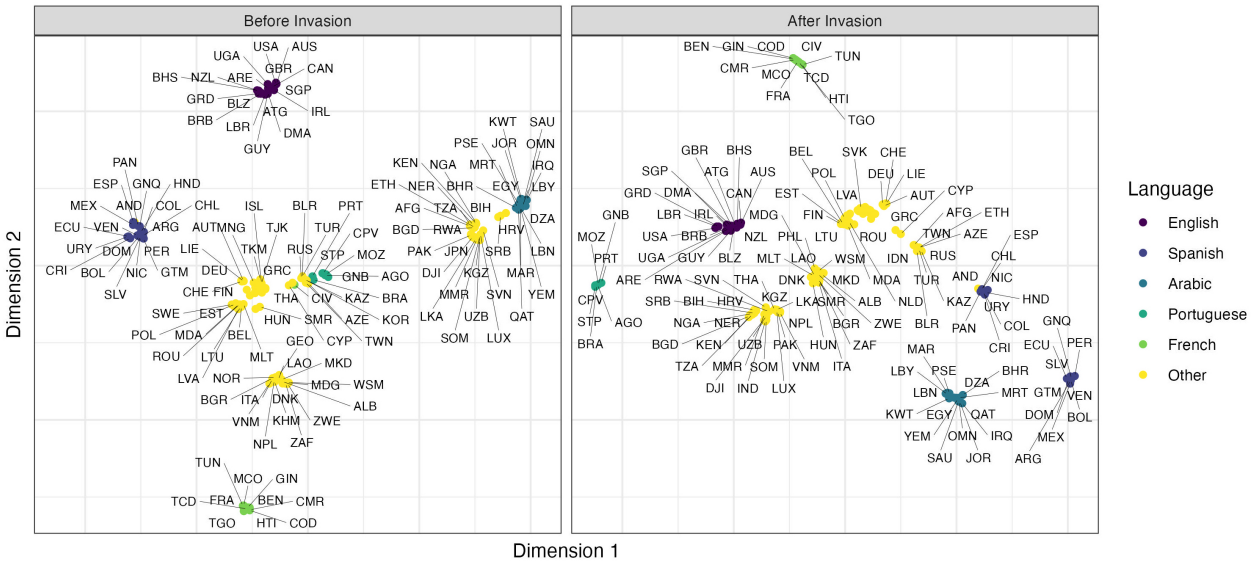
Figures 3.5 and 3.6 present visualizations of the similarity in shared sources between countries for the queries "Russia invasion" and "Ukraine crisis." For each query, we construct two N x N source similarity matrices, corresponding to our two rounds of searches. We populate the matrices with the mean proportion of shared sources across all queries for each dyad. We reduce the dimensionality of these matrices to N x 2 using Uniform Manifold Approximation and Projection (UMAP), facilitating visualization of the results. [75]

Figure 3.5: UMAP Visualization of Source Similarity: "Russia Invasion" Query



[75](McInnes, Healy, and Melville 2020) The reason we prefer this approach—constructing an N x N matrix of dyadic shared sources—over simply reducing the original source vectors is that it allows us to work with less sparse vectors, resulting in a dimensionality reduction process that offers more stability and less noise. Although the proximity of two states in this visual representation does not directly indicate a similarity in their sources, we can assume a degree of similarity when they are both similar to the same sets of other states. If two states share similar sources with other states, it is reasonable to infer that they also have some degree of source similarity between them.

Figure 3.6: UMAP Visualization of Source Similarity: “Ukraine Crisis” Query



Both before and after the invasion, we observe distinct clusters of countries based on language. These clusters encompass Arabic, English, French, Portuguese, and Spanish-speaking countries. Notably, for the “Russia invasion” query, there is a fragmentation of English-speaking countries into two clusters during both search waves. These two clusters are generally differentiated by geography, with one cluster primarily consisting of English-speaking countries in Africa and the Caribbean, and the other with countries in North America, Europe, and Oceania. An assessment of clustering performance in Appendix A.2 reveals more pronounced clustering after the invasion, suggesting heightened localization. By and large, however, our analysis predominantly indicates that countries within the largest language groups maintain consistent clustering behavior both prior to and following the Russian invasion of Ukraine.

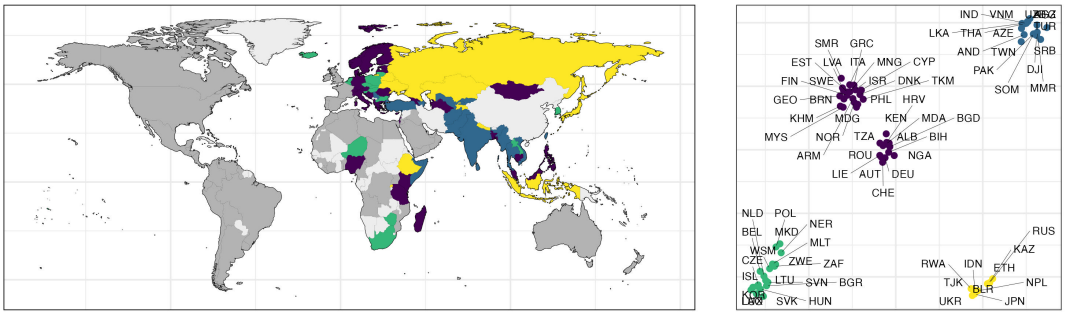
However, it is worth noting that there is indeed variation in the clustering behavior of countries beyond those belonging to the Arabic, English, French, Portuguese, and Spanish-speaking groups, observed across our two rounds of searches. Across all panels, we observe distinct clusters comprising countries from other language groups. We further investigate

this observation in Figure 3.7, where we exclude the previously mentioned language groups and focus solely on the remaining 78 countries. Building upon the UMAP results for this specific subset of countries, we perform hierarchical agglomerative clustering to automatically assign the countries into clusters. We determine the optimum number of clusters by using the maximum silhouette width as the criterion for any cluster count up to 20.[76] This process indicates four as the ideal number of clusters for all panels, except for the pre-invasion results for the "Ukraine crisis" query, where the optimal number is three. Figure 3.7 showcases the clustering outcomes for each round of searches, accompanied by a geographic representation of the clusters.

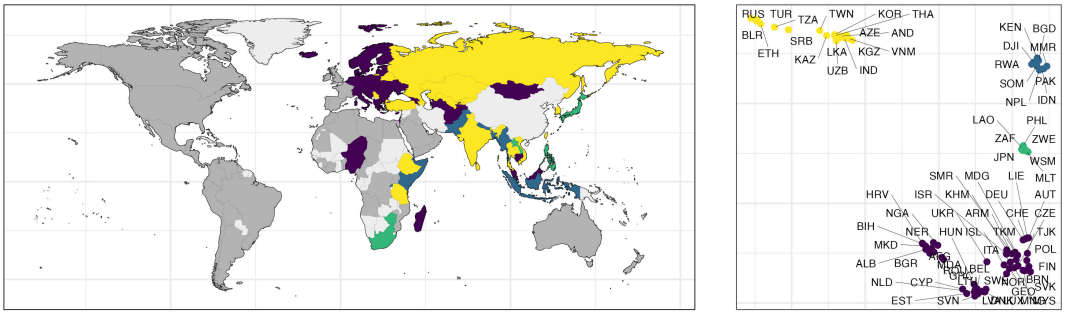
[76]We exclude Andorra from the "Ukraine crisis" results due to its behavior aligning with that of Spanish-speaking countries.

Figure 3.7: Clustering of UMAP Results from Smaller Language Groups

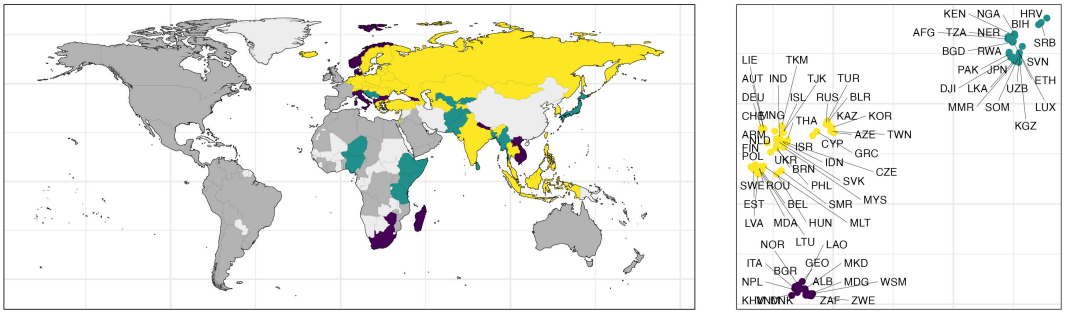
"Russia Invasion" - Before Invasion



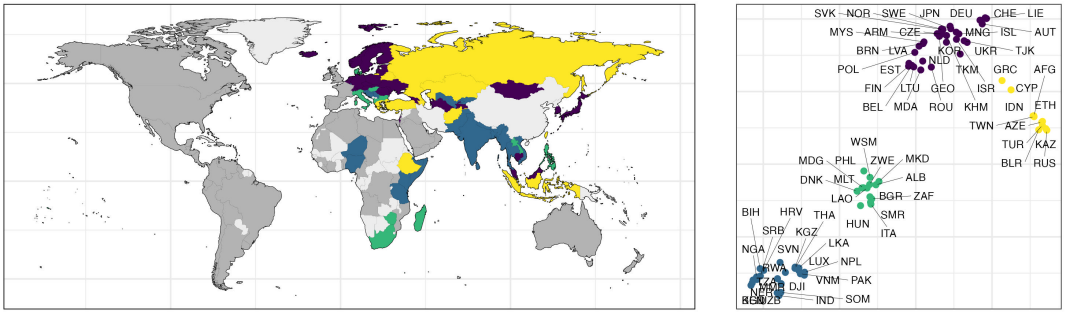
"Russia Invasion" - After Invasion



"Ukraine Crisis" - Before Invasion



"Ukraine Crisis" - After Invasion



Cluster 1 2 3 4 Major Language

For the "Russia invasion" query prior to the invasion, European states formed two distinct clusters, while Ukraine clustered together with Russia, Belarus, and Kazakhstan in a third group. Following the invasion, a significant shift occurred wherein nearly all European states, including Ukraine, converged into a single cohesive cluster. This suggests the emergence of a Eurocentric information pool, where European countries displayed increased exposure to similar publishers following the invasion of Ukraine. We find a similar trend in the results of the "Ukraine crisis" queries. Before the invasion, most European states clustered together with Russia, signifying a narrower disparity in accessible information for users within Europe and Russia concerning this issue. However, following the invasion, Russia formed a separate cluster with various non-European states, while a majority of European countries grouped together. The again suggests the emergence of a shared information pool among European countries post-invasion, and notably one that transcended language barriers.

While SERPs from the major language groups exhibit limited change following to the invasion, the observed variation in these less common language groups suggests that their information pools are more influenced by salient events. This may indicate the presence of "data voids" for political topics in less common languages, characterized by a scarcity of relevant content for specific search queries.[77] Within data voids, search results are more susceptible to change following significant events, given the limited competition among publishers to attain higher rankings on search engines. In summary, our findings suggest that countries within major language groups generally maintain their existing information pools after salient events, without transitioning to new non-language-based clusters. In contrast, countries with less common languages are more likely to experience changes in their information pools following salient events.

[77] Golebiewski and Boyd 2019.

3.6 Conclusion

Among scholars and policymakers, the Internet has long been envisioned as an equalizer, offering unprecedented opportunities to access and participate in the political sphere irrespective of nationality or geographic origin. By distributing massive amounts of information across billions of people, many expected the Internet to foster a more open, connected, and democratic world. However, this hopeful image is tarnished in a current landscape that is plagued by increasing concerns over misinformation, echo chambers, and digital censorship. Where existing IR scholarship contributing to this mounting skepticism has primarily concentrated on the role of state actors in limiting digital information flows, this chapter sheds light on the biases and skews propagated by search engines.

When individuals seek out information about international affairs, they do not encounter uniform search results. On the contrary, increasing personalization renders specific information more accessible to some than others, contingent on factors such as geographic origin and search language. As a result, different parts of the world inhabit different information pools, giving rise to diverse perceptions of international life.

Auditing approaches, like the one we conduct in this study, have major limitations and challenges. In addition to the prohibitive challenges to causal inference, real-world searches are affected by factors outside our purview, such as the query terms used, logged-in status, personalized settings, search history, and A/B testing conducted by the search engine itself. The algorithms underlying search engines are constantly changing, thus the insights we present in this chapter, and those presented in similar studies, offer a snapshot of search engine behavior from algorithms that may have already been superseded. Moreover, different demographic groups may seek out information in distinct manners, for instance by employing diverse search queries.[78] Although our methodology broadly approximates general search

[78] I. Weber and Jaimes 2011; Trielli and Diakopoulos 2022; Tripodi 2018.

behavior, it may not fully replicate the search patterns of specific demographic groups.

This chapter holds theoretical implications for several literatures in IR. First, our findings shed light on the unequal distribution of ideological and propagandistic content, contributing to the literature on psychological operations. These findings underscore that the Internet is not an equalizer, but rather an uneven playing field in which certain types of content are privileged over others. We observe a strong correlation between the dissemination of ideological and propagandistic content and search language, possibly driven by language-specific content production by publishers. Additionally, we find that countries with less widely spoken languages experience greater variation in their SERPs following significant international events, potentially due to the presence of data voids. These insights imply that publishers could strategically exploit data voids within the information pools of less common languages to effectively transmit ideological or propagandistic content.

Second, this research contributes to theories of algorithmic governance, which address the relationship between automated decision-making and the social organization and governance of society.[79] Our study highlights the significant role of algorithmic platforms like Google in the global dissemination of information. The availability of information that users encounter during their online searches holds deep-seated consequences for the formation of their world-views, especially when it comes to political attitudes and behaviors. Our findings highlight the influential role of personalization, linked to search language, location, and other factors, in shaping how Google portrays international affairs.

Lastly, this research has important implications for the study of norm diffusion. Our findings shed light on the existence of a distinct "language barrier" when it comes to accessing political information, including discourse on human rights and specific information related to major international conflicts. Different information pools encompass varied sets of resources concerning international issues and events, with the norms of content publishers embedded in

[79] Just and Latzer 2017; S. Srivastava 2021.

this information. The role of natural language in this context has received limited attention in the existing literature, yet our findings highlight its considerable influence as a mediator of information accessibility. These findings are significant as they suggest that individuals in different parts of the world may develop distinct understandings of international affairs in part because of their spoken language.

Search engines play a pivotal role as global information intermediaries, transcending the mechanisms of direct state control of internet censorship more rigorously studied in IR. This research highlights the complexities of localized search results for queries related to international politics. By way of the various roles that publics play in international politics, these findings suggest that search engines may increasingly shape international political dynamics. This becomes ever more significant as search engines continue to expand their reach to every corner of the globe.

Chapter 4

Keep Your Enemies Closer: Word Embedding Representations of Friendship and Collective Identity Between States

4.1 Introduction

Policymakers and the public frequently use the languages of friendship and collective identity to describe international politics. Political leaders often refer to nations as friends, such as former Mexican President Felipe Calderón's statement that "Mexico is a good friend of Cuba, and Mexico is also a good friend of the United States," and the U.S. Department of State's declaration that "Israel has no greater friend than the United States."[1] The public understands this language as well, where a growing percentage of Americans view China as an "enemy" of the United States.[2] Policymakers and the public also refer to collective identity among nations, as seen in European Council President Charles Michel's statement "We are all Ukrainian" during a recent visit to the country.[3]

While collective identity is more rigorously theorized in the field of international relations, friendship is often reduced to a rhetorical device in the description of international politics. Furthermore, although both concepts describe some sort of affinity that states have for each other, it is unclear how they might relate to the concept of "geopolitical affinity," which is common in quantitative studies. Geopolitical affinity is often quantified through measures such as United Nations General Assembly (UNGA) voting behavior, shared membership in international institutions, or formal alliances, but it remains ambiguous whether these formal proxies describe friendship, a shared identity, or something else entirely. Moreover, quantitative studies of friendship and collective identity in international relations are rare, perhaps unsurprisingly so given the challenges in measuring concepts that theoretically cannot be reduced to material factors. As Jervis (1997) warns us, "one cannot infer results from desires and expectations and vice versa"—friendship and collective identity are

[1] *Joint Press Conference With President Barack Obama And President Felipe Calderon Of Mexico* 2009; *U.S. Relations With Israel* 2021.

[2] (Frankovic and Orth 2023) 43% of Americans view China as an "enemy" in 2023, up from 9% in 2018.

[3] Camut and Boonen Februaruy 23, 2023.

concepts rooted in actor perception, and it is therefore challenging to infer them from any observable outcome.[4] Measuring friendship and collective identity with formal proxies has inherent limitations, not the least of which is understanding how these concepts relate to each other.

This chapter develops previous research on neural network embedding models to represent the perception of friendship and collective identity between countries as relationships between vectors in high-dimensional space. Using all floor speeches from the 56th to the 114th sessions of the Congress of the United States, I use word embedding models to create biennial friendship and collective identity measures between the United States and 192 countries over a period of 118 years. Word embedding models are trained on large quantities of text to produce representations of words as vectors based upon the context in which a word is used. I extend the methodology introduced by Kozlowski et al. (2019) in which "dimensions" created by antonym pairs in an embedding space—such as *friend–enemy* or *us–them*—are used to evaluate the meaning of a set of target words.[5] By projecting the word vectors of country names along a *friend–enemy* dimension, for example, countries typically considered as friends of the United States, such as Canada, tend to appear on one end of the dimension, while countries such as Russia and Iran tend to appear on the other. Importantly, these measures are based on the perception of speakers—they describe the context in which speakers refer to countries and do not solely depend on overt mentions of friendship or identity. With this data set, I assess two primary questions. First, what are the factors that explain the extent of perceptions of friendship and collective identity between states? Second, as components of geopolitical affinity, how do friendship and collective identity relate to each other?

I find that perceptions of friendship between the United States and other countries grew dramatically during and immediately following World War II. The factors that contribute

[4]Jervis 1997, p. 578.

[5]Kozlowski, Taddy, and J. A. Evans 2019.

to friendship have remained largely consistent over time, with countries that share similar political systems and trade relationships having a positive association with one another. In contrast, the United States has experienced a steady strengthening of perceptions of collective identity with foreign states over the last 80 years, but the factors which explain collective identity have varied over time. In recent decades, the correlates of enmity—including material strength and a history of conflict—have also come to predict collective identity. This phenomenon suggests the development of "transactional" friendships, in which friends may not necessarily share a collective identity, as well as a deeper paradox: the United States may share a sense of identity with states it considers to be enemies, indicating the existence of a collective identity amongst states engaged in a certain level of competition.

This research contributes to the understanding of geopolitical affinity by explicitly analyzing its components and correlates. While past studies which use formal proxies for geopolitical affinity have yielded important insights in the field, their methods have lacked the precision necessary to differentiate between distinct types of affinity between states. In contrast, my methodology goes beyond formal or even rhetorical measures by analyzing geopolitical affinity in a discursive manner. Specifically, I examine the context in which legislators discuss a state and compare it to the context in which they discuss friends, enemies, themselves, and others. Furthermore, this research challenges the notion that it is necessary to "other" another group to view them as an enemy.[6] I demonstrate that perceiving another state as an enemy is not necessarily synonymous with othering them. Finally, this chapter represents a contribution to the applications of word embedding models by extracting meaning from rare words in a corpus using a dimensional approach.

[6]Mearsheimer 2018.

4.2 Friendship and Collective Identity in International Relations

I conceptualize friendship and collective identity in a Bourdieusian sense, in which they are practices carried out by social actors. Bourdieu writes that an actor's sense of place and habitus are "at the basis of all forms of cooptation [and] friendships", regulating their social practices.[7] In the case of friendship, for example, an actor will demonstrate friendship-related practices towards others that they perceive as their friends. In this chapter, I specifically examine speech, a core practice through which actors participate in social interactions.

Compared to enmity, the notion of friendship is substantially undertheorized in IR literature. Wendt (1999) points out that this is with good reason—hostilities between states are of greater consequence to international politics, and to extend the notion of friendship to states seems to push the boundaries of anthropomorphism.[8] Nevertheless, the language of friendship pervades international discourse: the UN Charter emphasizes the promotion of "friendly relations among nations", the United States has cultivated a "special relationship" with the United Kingdom and refers to France as "America's oldest friend", and many "friendship treaties" exist between states.[9]

Theories of friendship date to Greek philosophy. Aristotle, for example, developed a typology of three types of friendship.[10] The first is utility friendship, where actors are friends due to mutual benefit; the second is pleasure friendship, where love is directed towards the pleasure that the other offers, rather than towards the person themselves; the third is virtue friendship, which is based on mutual recognition of good character and involves loving the other person as a whole. In more recent times, Carl Schmitt describes the friend–enemy

[7] Bourdieu 1989, p. 17.
[8] Wendt 1999, pp. 298–299.
[9] *The United States and France: Allies, Partners, and Friends* 2021.
[10] Aristotle and Irwin 1999.

distinction as the specific notion to which political motives and actions can be reduced.[11] Schmitt also identifies two types of friendship—utilitarian and existential friendship—which generally align with Aristotle's utility and virtue friendship, respectively.

Wendt posits that friendship between states entails a role structure in which states expect their friends to follow two rules: 1) disputes will be settled without war or the threat of war, and 2) friends will fight as a team if either is threatened by a third party.[12] It is in this vein that scholars have identified friendship as underlying both pluralistic securities communities—in which members settle disputes amongst themselves without violence—and collective security arrangements, in which states fight together against third parties.[13] Any friendship that might underpin a formal alliance is best associated with utility friendship, in which states "feel individually threatened by the same threat" and the collaboration ceases when the common threat is gone.[14] The friendship underlying more complex arrangements, especially collective security arrangements, approaches virtue friendship to the extent that actors collectively identify with the fate of others in the arrangement and band together for more than instrumental purposes.

Collective identity among states is shaped by a variety of factors, including racial, historical, political, and cultural elements.[15] Role-playing in international institutions can also contribute to the formation of collective identity.[16] According to Wendt (1994), collective identity emerges endogenously among states at the system level, which can lead to cooperation.[17] Furthermore, collective identity is a key factor in the formation and maintenance of security communities. Collective identification plays a constitutive role in security com-

[11] Schmitt 2007.
[12] Wendt 1999.
[13] Deutsch 1957; Wendt 1994.
[14] Wendt 1999, p. 301.
[15] Hemmer and Katzenstein 2002.
[16] Checkel 2005.
[17] Wendt 1994.

munities by redefining state interests, which are socially constructed through interaction.[18] Additionally, security communities reinforce collective identity by relying on mutual trust and a sense of "we-ness" to facilitate peaceful dispute resolution.[19] Taken together, participation in security communities can strengthen collective identity, but it is also instrumental in enabling their creation.

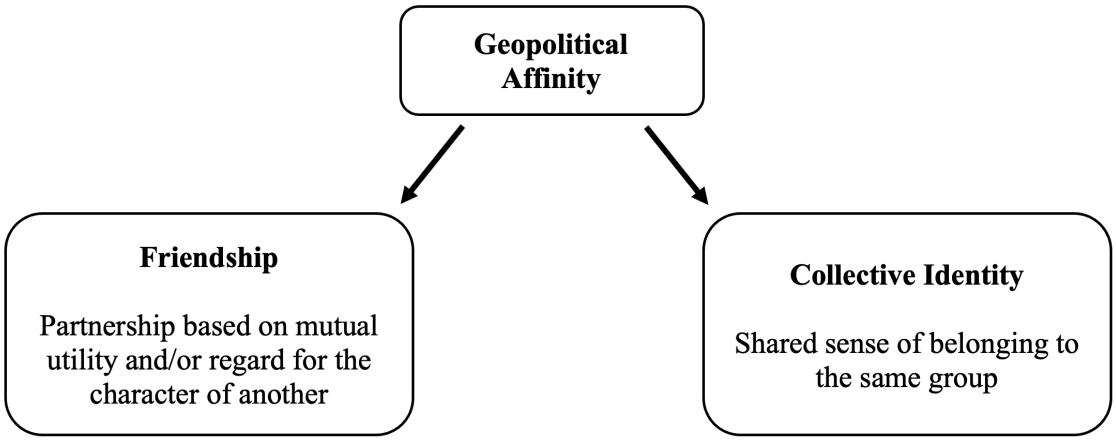
As collective identity relates to friendship, Weber (1978) developed the idea of social closure—or the openness of social relationships to outsiders—as being indifferent towards friendship or approval.[20] While closed social relationships might be associated with warm feelings towards other members, such as in personal emotional relationships, they may also be characterized by competition for resources among members, for instance in the case of stock exchange brokers. As it relates to states in the international system, the relationship between collective identity and friendship remains unclear.

Empirical measures of friendship and collective identity in IR literature are sparse. Quantitative studies more often capture "geopolitical affinity", which is usually proxied through an observable display of foreign policy preference, most commonly UNGA voting behavior.[21] Many studies, however, criticize the use of UN voting to measure geopolitical affinity, with one study reminding us that such a measure "is not an expression of closeness of relations between two states but an expression that both states agree on the desired fate of a UN resolution."[22] Other studies note that UN voting is strategic and that "vote buying" in the UNGA occurs through foreign aid dispersals.[23] For many reasons, vote outcomes may not directly reflect the extent to which states consider each other to be friends or collectively identify with one another.

[18] Adler and Barnett 1998; Pouliot 2007, p. 14.
[19] Deutsch 1957.
[20] Duque 2018; M. Weber, Roth, and Wittich 1978, p. 580.
[21] I. T.-y. Chen 2022; Copelovitch and Rickard 2021; Terman and Byun 2022; Zarpli and Zengin 2022.
[22] Voeten 2021, p. 106.
[23] Alexander and Rooney 2019; D. B. Carter and Stone 2015; Terman and Byun 2022.

Figure 4.1 depicts the connection between friendship, collective identity, and geopolitical affinity, as conceptualized in this chapter. It is important to emphasize that although I consider friendship and collective identity as interconnected, they are distinct concepts, each describing a unique form of affinity that exists between states.

Figure 4.1: Dimensions of Geopolitical Affinity



4.3 The Dimensional Approach to Word Embeddings

Word embedding models analyze a large text corpus and use it to deduce the meanings of words, which are then expressed as vectors. These models rely on the distributional hypothesis that the context in which a word is used is reflective of its meaning.[24] It follows that the meaning of a target word can be inferred from the neighboring words in a text; moreover, if two words tend to occur around similar sets of words, they likely have similar meanings. Consider the following sentences:

- My friends played *basketball* outside after school.
- My friends played *football* outside after school.

[24]Z. S. Harris 1954.

A word embedding model would recognize the similarities between the neighboring words of *basketball* and *football*, and thus ascribe to them more similar vector representations. Once an entire embedding space is constructed, words with similar meanings—based upon the context in which they are used in the corpus—are positioned close to one another.

I use fastText to construct word embeddings.[25] FastText differs from other word embedding algorithms, like Word2Vec, in that it generates embeddings for sub-word character strings instead of individual words.[26] To produce an embedding for a word, the algorithm sums the embeddings of the sub-word character strings. Consider the example of "Canada": to construct the word vector, fastText would sum the embeddings of "ca", "can", "ana", "nad", "ada", and "da".[27]

This approach offers several advantages over other common word embedding algorithms. FastText is generally effective in constructing meaningful embeddings for rare words in a corpus. By virtue of their shared sub-word character strings, words such as "Canadian" or "Canadians" help construct the embedding vector of "Canada." This feature is particularly relevant for the analysis in this chapter since most country names are infrequently mentioned by legislators in Congress. Another advantage of using fastText is its ability to handle transcription errors. By embedding subword character strings, fastText can still infer meaningful information even if a word is transcribed incorrectly. For instance, if "Canada" is transcribed as "Canad," fastText still learns the context for the substrings "ca," "can," "ana," and "nad," which contribute to the embedding for "Canada." In contrast, other word embedding algorithms that embed entire words would treat "Canad" as a unique word and not infer any information about the meaning of "Canada."[28] The main tradeoff in using

[25]Bojanowski et al. 2016.

[26]Mikolov, K. Chen, et al. 2013.

[27]In practice, the analyst sets a minimum and maximum length of character strings for which to learn vector representations, and fastText sums all character strings of matching lengths to produce a word vector. In this chapter, I set the minimum sequence length to 3 and the maximum length to 6.

[28]This feature is particularly pertinent for this analysis because the corpus contains many transcription errors (see Section 4.4).

fastText over algorithms that embed whole words, such as Word2Vec, is that these other algorithms are known for generally positive performance on semantic similarity and analogical tasks, which are the primary goals of this chapter.[29] However, because my principal concern is investigating rare words in a corpus that contains a significant number of transcription errors, I opt to use fastText to optimize the stability of the rare word embeddings.

Previous research has shown that it is not only the distance between vectors in an embedding space that describe differences in word meaning, but also the direction of the distance.[30] A common exercise is to perform mathematical analogies in the embedding space. For example, if an analyst were to begin with the embedding of the word "king", subtract the embedding of the word "man", and add the embedding of the word "woman", it is likely that the resulting vector would be closest to the embedding of the word "queen."[31] In other words, the direction of the distance between the words "king" and "queen" can be described as a direction of gender, analogous to the distance between the words "man" and "woman."

Following this intuition, the "dimensional" approach to word embeddings involves creating "dimensions" in the embedding space to represent social phenomena of interest. The dimensional approach was initially introduced by Kozlowski et al. (2019) to measure markers of social class.[32] Since its introduction, other scholars have used variations of this method to assess intersectionality in the U.S. South in the 19th Century, political polarization in online communities, and gender attitudes among judges in U.S. circuit courts.[33]

In the dimensional approach, the analyst identifies antonym pairs that serve as the "poles" of each dimension—in the case of gender, the analyst might use antonym pairs such as *man–woman* or *masculine–feminine*. The optimal choice of antonym pairs are words which differ in meaning along the target dimension but are otherwise similar. The analyst then

[29]Dawn Chen, Peterson, and Griffiths 2017; Mikolov, Sutskever, et al. 2013.
[30]Mikolov, K. Chen, et al. 2013.
[31]Ibid.
[32]Kozlowski, Taddy, and J. A. Evans 2019.
[33]Ash, Daniel Chen, and Ornaghi 2021; Nelson 2021; Waller and A. Anderson 2021.

calculates a single vector $\boldsymbol{d}$ to represent the dimension by averaging the distance between the word vectors of the antonym pairs. Using this vector $\boldsymbol{d}$, the analyst can compute the score of other words in the corpus along the dimension by taking the orthogonal projection of a word vector $\boldsymbol{w}$ onto the dimension vector $\boldsymbol{d}$. The resulting score represents the proportion of similarity to one pole of the dimension minus similarity to the other pole. The further a word's score deviates from 0, the more the meaning of the word is associated with one pole of the dimension.

4.4 Dimensions of Geopolitical Affinity in the Speech of Legislators

The text source for this study is the text of the United States Congressional Record from the 56th to the 114th Congress, covering the years 1899-2017. This corpus is a "substantially verbatim" record of all daily floor proceedings in the United States House and Senate.[34] I access the corpus through Stanford's Social Science Data Collection, which converted the original documents into plain text format using OCR.[35]

This corpus offers numerous advantages, including its extended time span and diverse political perspectives. However, there are also several limitations. By selecting this corpus, I assume that perceptions of friendship and collective identity between the United States and other countries are conveyed through the speech of legislators. However, these views may not necessarily mirror the perceptions of the American public, political leaders, or bureaucrats. While I do not address these other actors within this chapter, future research may consider leveraging corpora from social media, newspapers and other media outlets, or presidential briefings and speeches.

[34] Amer 1993.
[35] Gentzkow, Shapiro, and Taddy 2019.

Additionally, some words that are ideal selections for antonym pairs are used in unique ways by legislators. For example, the word "friend" is often used to refer to other legislators (e.g., "My friend from Illinois"), which may differ from its meaning in everyday discourse. To address this issue, I conduct robustness checks in Appendix B.1 using dimensions constructed without these unique words. The results are largely consistent with the main findings.

Using the dimensional approach to word embeddings, I represent two notions of geopolitical affinity—friendship and collective identity—as relationships between vectors in high-dimensional space. I construct a dimension vector $\boldsymbol{d}$ with each concept's associated antonym pairs. For the friendship dimension, I begin with a starter set of antonyms (friend–enemy, ally–adversary) and consult the Merriam-Webster Thesaurus to generate a complete list of synonyms for each pole. As Kozlowski et al. (2019) note, this procedure requires some discretion on the part of the analyst, as thesauri "often contain a wide range of loosely synonymous terms inappropriate for the given analysis."[36] The curated set of synonyms for each pole includes those referenced in the thesaurus that appear to be contextually relevant for political speech.[37]

I follow a slightly different procedure for the collective identity dimension. Here, I use a list of the most common possessive (ours–theirs), personal (us–them), and reflexive (ourselves–themselves) pronouns in the English language, with first person pronouns representing the "us" pole of the dimension and third person pronouns representing the "them" pole. Previous research finds that although additional antonym pairs generally lead to more robust results, there are substantially diminishing returns as more antonym pairs are added.[38] I present an additional analysis in Appendix B.1 using only the starter sets of antonyms for each dimension, and the full list of antonym pairs used to construct each dimension is presented in Appendix B.2.

[36]Kozlowski, Taddy, and J. A. Evans 2019, p. 919.
[37]For example, for the "friend" pole, I exclude synonyms such as "mate" and "brother".
[38]Kozlowski, Taddy, and J. A. Evans 2019.

To prepare the data for analysis, I divide the corpus into 59 sub-corpora based on each session of Congress, which lasts approximately two years. I preprocess the text by removing capitalization and punctuation, and I treat each utterance by a successive speaker as a separate document. I also standardize the names of countries which have undergone a name change or are referred to with multiple names, replacing all mentions of a country with a consistent name. For example, I replace "USSR", "Soviet Union", "United Socialist Soviet Republic", and "Russia" with "russia". Additionally, I follow the fastText documentation by concatenating all country names that are not single words, such as "Costa Rica" being transformed to "costa_rica". I provide a list of all instances of country preprocessing in Appendix B.3.

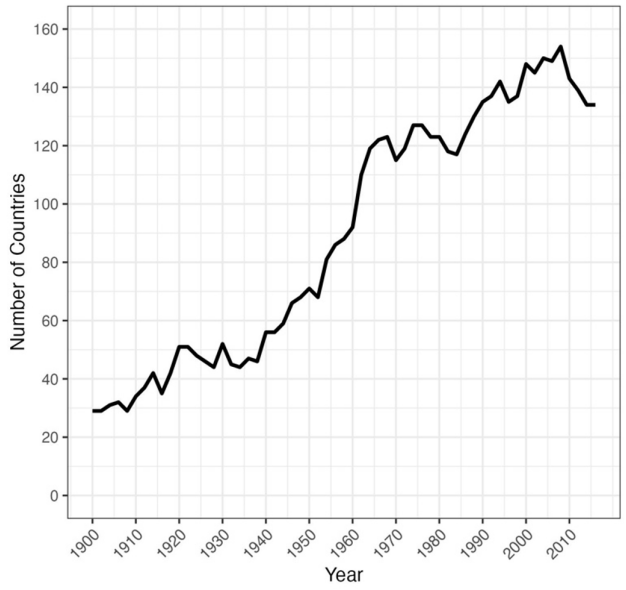
I create separate word embedding models for each session of Congress and use nonparametric bootstrapping to assess the uncertainty in my estimates. To create 90% bootstrapped confidence intervals, I randomly sample documents with replacement from each session's corpus to create 20 new sub-corpora of equal size. For each sub-corpus, I produce a separate word embedding model, construct dimension vectors for friendship and collective identity using antonym pairs, and calculate the orthogonal projection of each country word vector $\boldsymbol{w}$ on the dimensions. I then take the second lowest and second highest projection on each dimension from the 20 models to serve as the 90% confidence bound.[39] I set the minimum vocabulary frequency for the models to 10, meaning that a country name must appear at least 10 times in the sub-corpus to appear in the model. In the following analysis, I only consider country observations that appear in all 20 bootstrapped models for a given session of Congress. Finally, following the recommendations of Rodriguez and Spirling (2022) for minimum context window size and word vector length, I employ a context window of 5 and produce word vectors of length 300.[40]

[39]Bickel and Freedman 1981; Efron 2003; Singh 1981.

[40](Rodriguez and Spirling 2022) Rodriguez and Spirling (2022) recommend avoiding context windows under length 5. (Mikolov, Sutskever, et al. 2013) also show that word vectors of length 300 capture a more

Figure 4.2 showcases the count of observations generated per session of Congress, encompassing only country names that appear in all 20 bootstrapped samples. The range of observations per session of Congress spans from 29 countries in the early 20th century to a maximum of 154 in 2008. The variation in the overall country count is influenced by both the number of countries existing globally in a specific year and the total volume of speech during each congressional session (for additional information, refer to Appendix B.4). Beside the figure, I present a list of the countries for which there are the highest and lowest number of observations in the data set.

Figure 4.2: Observation Count per Session of Congress



Highest Frequency (Appear in All Sessions)
Australia, Belgium, Brazil, Canada, China, Colombia, Denmark, France, Germany, Italy, Japan, Mexico, The Netherlands, New Zealand, Nicaragua, Russia / Soviet Union, Spain, Sweden, Switzerland, Turkey, United Kingdom

Lowest Frequency (Count in Parentheses)
Andorra (0), Antigua and Barbuda (0), Comoros (0), Kiribati (0), Maldives (0), Seychelles (0), Tuvalu (0), Bhutan (1), Nauru (1), San Marino (1), Suriname (1), Vanuatu (1), Zanzibar (1), North Yemen (2), Papua New Guinea (2), Solomon Islands (2)

nuanced representation of word meaning than shorter length vectors.

4.5 Validating Dimensions of Geopolitical Affinity

4.5.1 Validation Exercises

I validate my methodological approach using a combination of ground truth data and practical exercises. For the friendship dimension, I validate the country projections by comparing them to survey data from the Chicago Council on Global Affairs. From 1978 through 2010, the Chicago Council's quadrennial American Public Opinion and U.S. Foreign Policy survey asked respondents to rate foreign states on a feeling thermometer from 0-100, with a focus on politically relevant states.[41] For example, the 2010 survey asked respondents to rate Great Britain, China, Brazil, India, Japan, Russia, and South Korea, as well as randomly soliciting a rating for about half of a longer list of 15 countries.

In this exercise, I make the assumption that the feeling thermometer scores from the Chicago Council survey can serve as a useful proxy for how legislators perceive friendship with foreign states, although it is likely an imperfect one. Moreover, the survey data provides a brief snapshot of public opinion during the two-year period on which each friendship projection is based—any variation between this snapshot and the mean feeling thermometer score for the two-year period would appear as noise in this correlation test.

Figure 4.3 displays the country feeling thermometer scores against the friendship projections for the year in which the survey data was collected. The correlation between the feeling thermometer scores and friendship projections is 0.566. It is evident from Figure 4.3 that states which appear on the extremes of either scale tend to exhibit stronger correlations than those in the middle. In fact, when considering only states whose feeling thermometer values are less than 40 or greater than 60, the correlation between feeling thermometer scores and friendship projections increases to 0.849. This suggests that the friendship projections are

[41] Foreign Relations 2021.

most effective in identifying states with whom legislators perceive either a strong friendship or a strong enmity.

Figure 4.3: Correlation of Feeling Thermometer and Friendship Projections

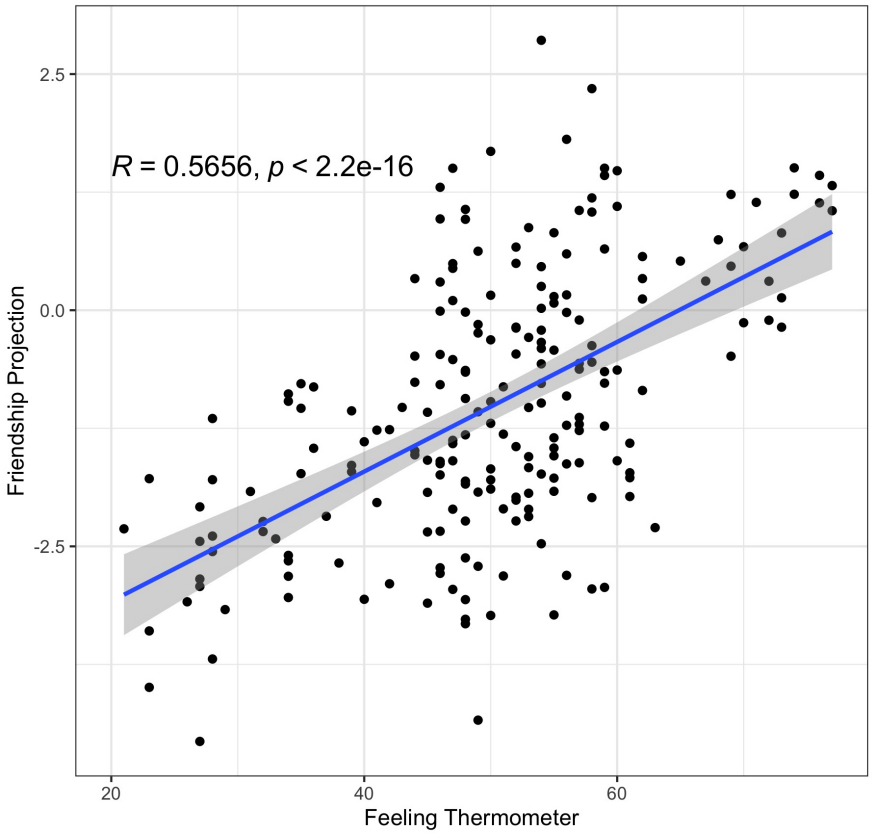


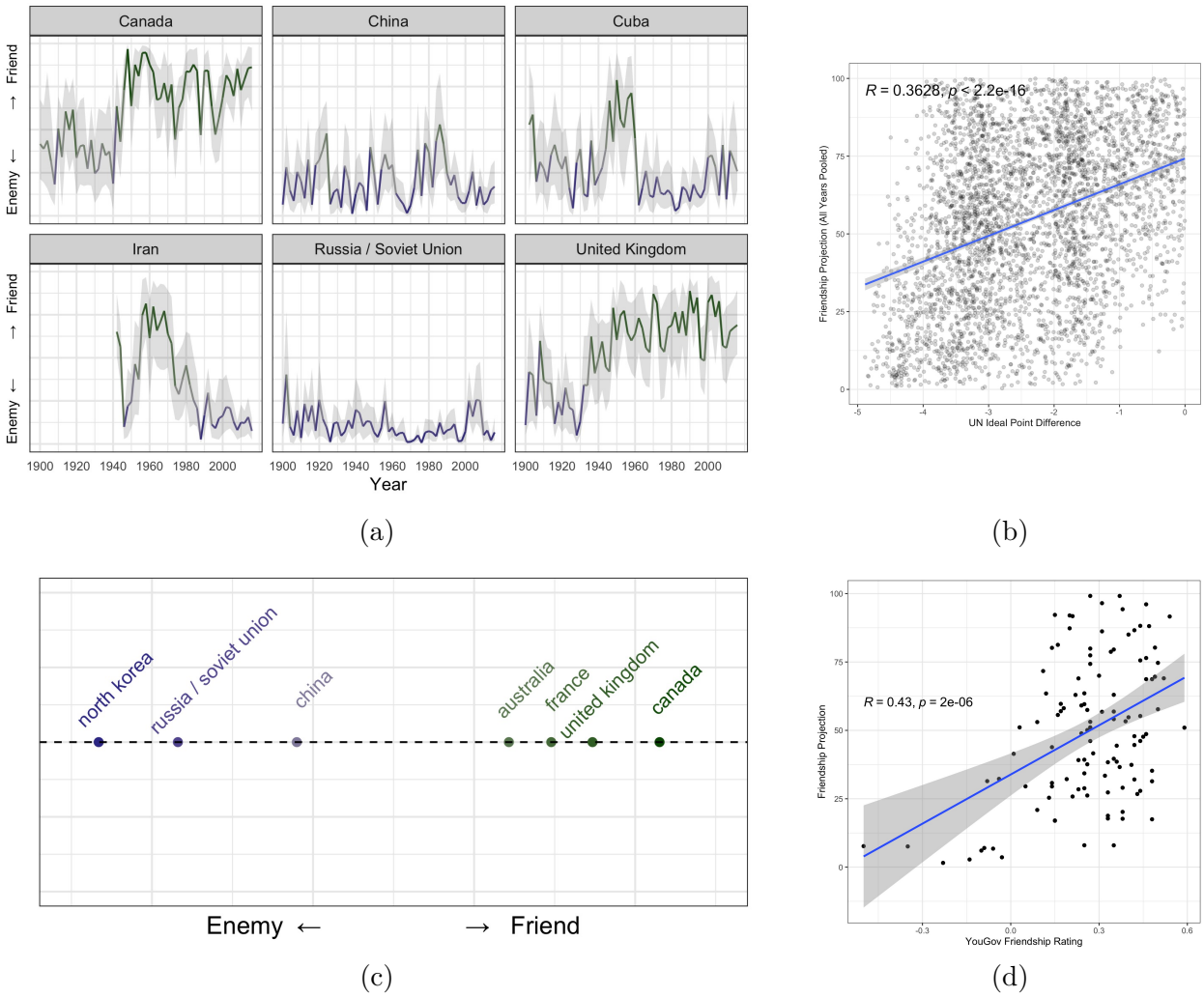
Figure 4.4 presents the results of several other exercises to assess the validity of the friendship projections. Figure 4.4(a) illustrates how the friendship projections capture defining moments in key bilateral relationships, such as the Cuban Missile Crisis and the severing of diplomatic ties with Cuba, Cold War tensions with the Soviet Union, and the Iranian Revolution. Figure 4.4(b) shows the average friendship projections for various geopolitically relevant states. Notably, states widely considered to be strong allies of the United States are projected onto the friend pole of the friendship dimension, while states perceived as adversaries are projected onto the enemy pole.

In Figure 4.4(c), I observe a moderate positive correlation of 0.363 between the friend-

ship projections and UN Ideal Point data, which measures geopolitical affinity through voting agreement in the UNGA. Figure 4.4(d) plots the correlation between the 2016 friendship projections and a 2017 public opinion survey on international friendship conducted by YouGov, where respondents from the American public rated certain states as friendly or unfriendly with the United States.[42] The correlation of 0.430 is lower than the Chicago Council survey, and similar challenges arise due to differences in the views of the public and policymakers, as well as the misaligned time horizons between the survey and projections. In Figure 5, I plot the distribution of all friendship projections across all sessions of Congress.

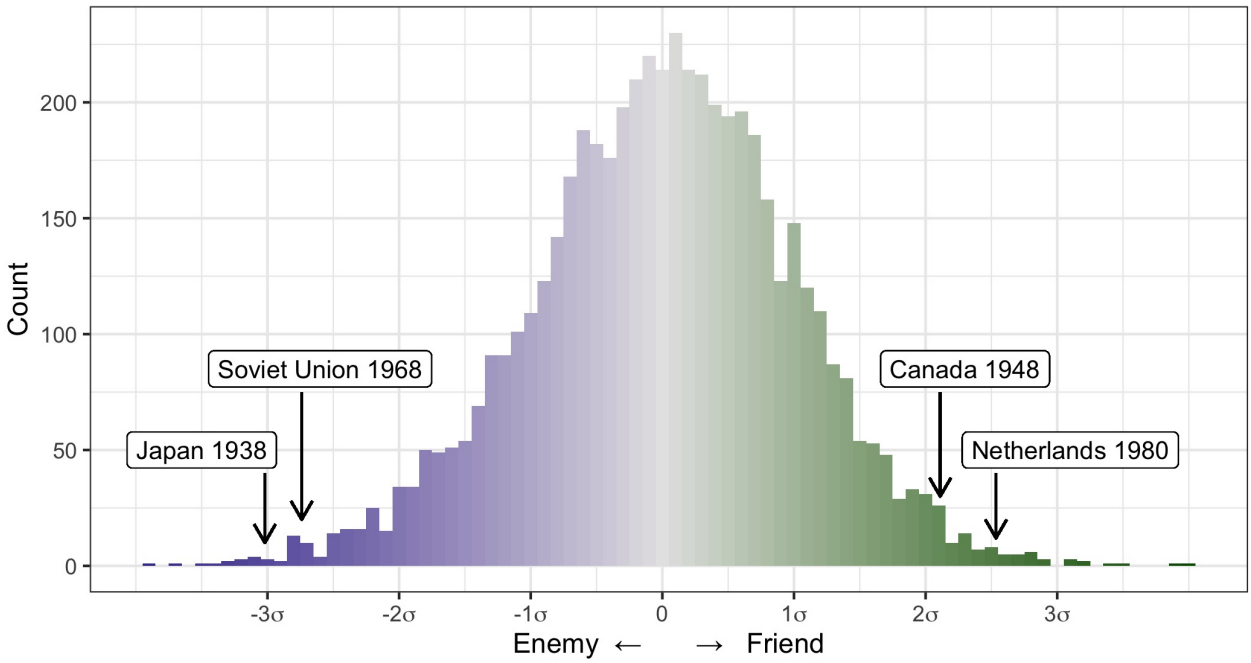
[42]"America's Friends and Enemies" 2017.

Figure 4.4: Validation Exercises for Friendship Projections



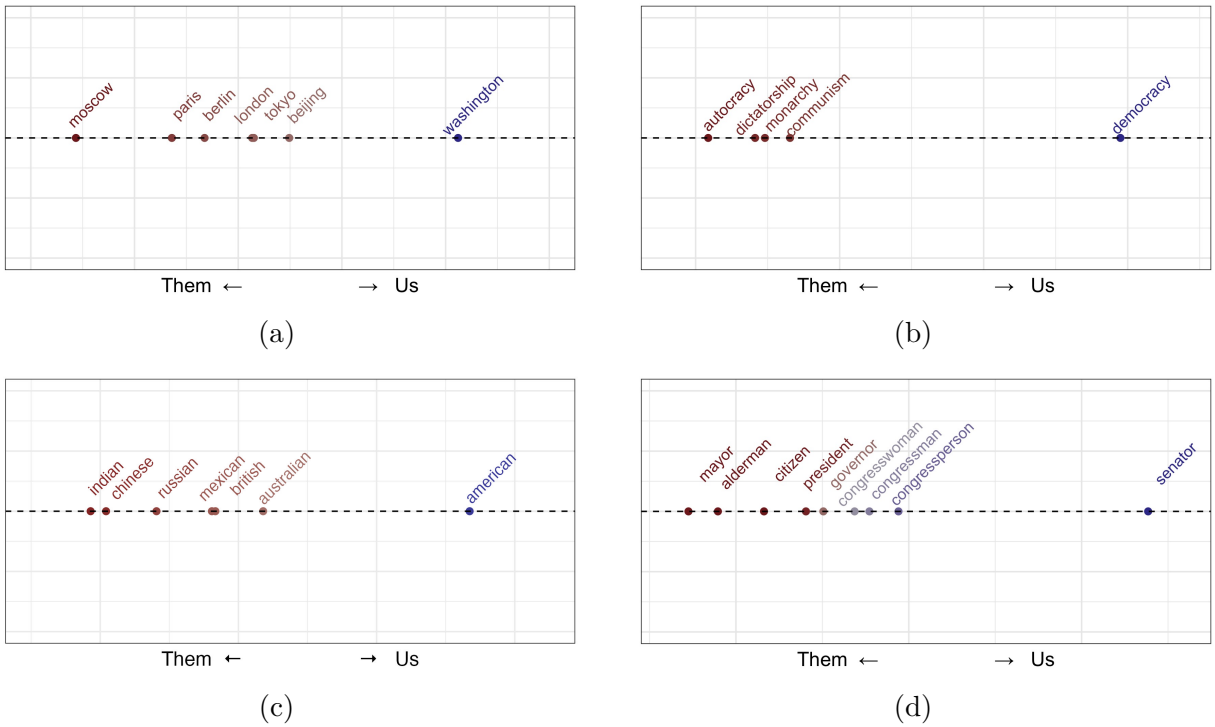
Validating the collective identity dimension is challenging because there is minimal survey data or other ground truth data available on the topic. To assess the validity of these projections, I examine a variety of topics where there is an "American" and a "non-American" alternative. For example, Panel (a) in Figure 4.6 plots the mean collective identity projections of a selection of capital cities across all years in the data set. I find that "Washington" is projected closest to the "us" pole of the dimension, while capital cities of other states are projected towards the "them" pole. I repeat this exercise for forms of government (Figure 4.6(b)) and nationalities (Figure 4.6(c)), finding that "democracy" and "American" are

Figure 4.5: Distribution of Friendship Projections Across All Years



projected closest to the "us" pole. In Figure 4.6(d), I repeat this exercise for various government roles. The speakers in the corpus are primarily senators and congresspeople, and I find that these roles are projected closest to the "us" pole, while other government roles such as governor and mayor are projected closer to the "them" pole.

Figure 4.6: Validation Exercises for Collective Identity Projections

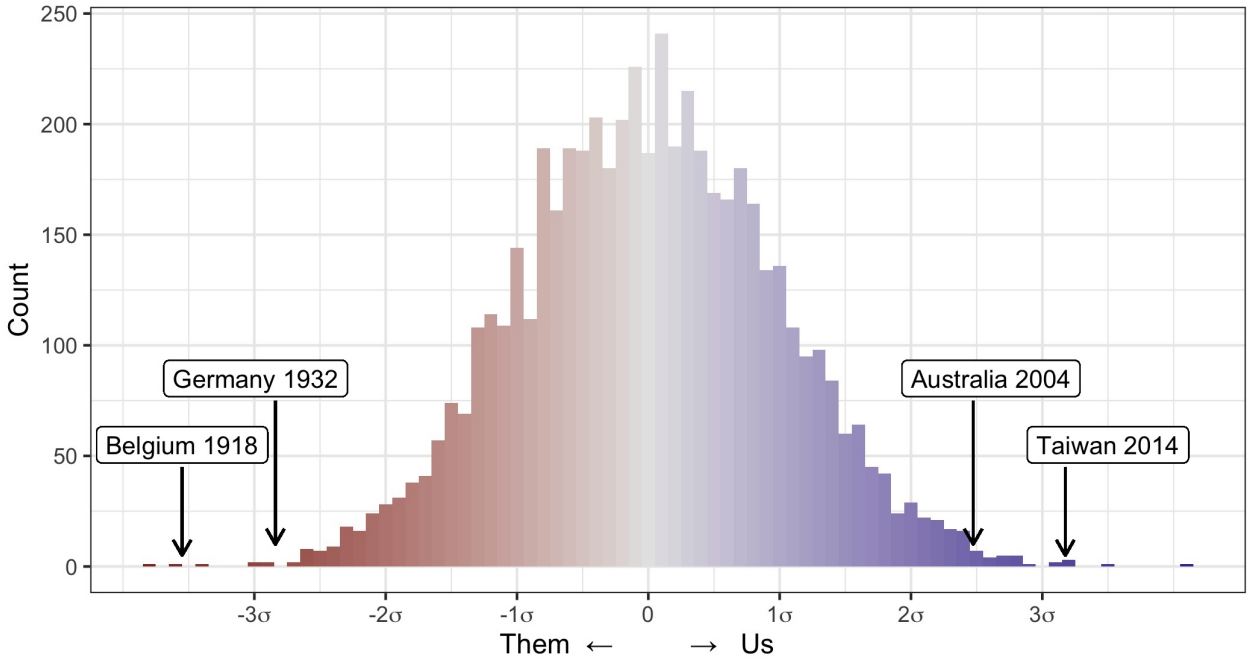


In addition to the challenges in validating the collective identity projections, there is likely less reliability in these projections due to the types of antonym pairs used to construct the dimension vector $\boldsymbol{d}$. While the antonym pairs used to construct the friendship dimension are directly related to friendship and enmity, using a set of pronouns to construct the collective identity dimension likely introduces greater variability due to the broader application of these words in various speech contexts. Consequently, there is likely greater noise associated with the construction of the dimension vector $\boldsymbol{d}$, resulting in less reliability in the collective identity projections as compared to the friendship projections.

In Figure 4.7, I display the distribution of all collective identity projections. Notably, the projection for Belgium in 1918 is one of the strongest "them" projections in the data set. Considering its significance as a site of World War I, it is plausible that many references to Belgium at that time revolved around its role as a battlefield, rather than specific attributes

of the country or its government. The collective identity projection for Belgium would shift towards the "them" pole when the descriptions of war events referencing "Belgium" diverge from those describing "us" or "our" country. Appendix B.5 includes placebo tests which further suggest that the collective identity projections are less stable than the friendship projections. Future research should conduct more extensive validations of collective identity projections to understand their strengths and weaknesses, ideally leveraging ground truth survey data on the subject.

Figure 4.7: Distribution of Collective Identity Projections Across All Years



Finally, I present the most common nearest neighbors for a selection of relevant words in the corpus. Drawing from the recommendation of Rodriguez and Spirling (2022), I curate a set of relevant terms and calculate their nearest neighbors in the embedding space for each session of Congress.[43] The curated words aim to showcase the dimensions of friendship and collective identity, countries of primary interest, and other pertinent topics that regularly

[43]Rodriguez and Spirling 2022.

arise in the speech of legislators. For each curated word, its nearest neighbors are the other words in the embedding space with the most similar vector representations. I calculate the 50 nearest neighbors for each term in each session of Congress and rank the results according to the total number of sessions in which they appear, presenting the top 10 results for each curated word in Table 4.1.

Table 4.1: Most Common Nearest Neighbors

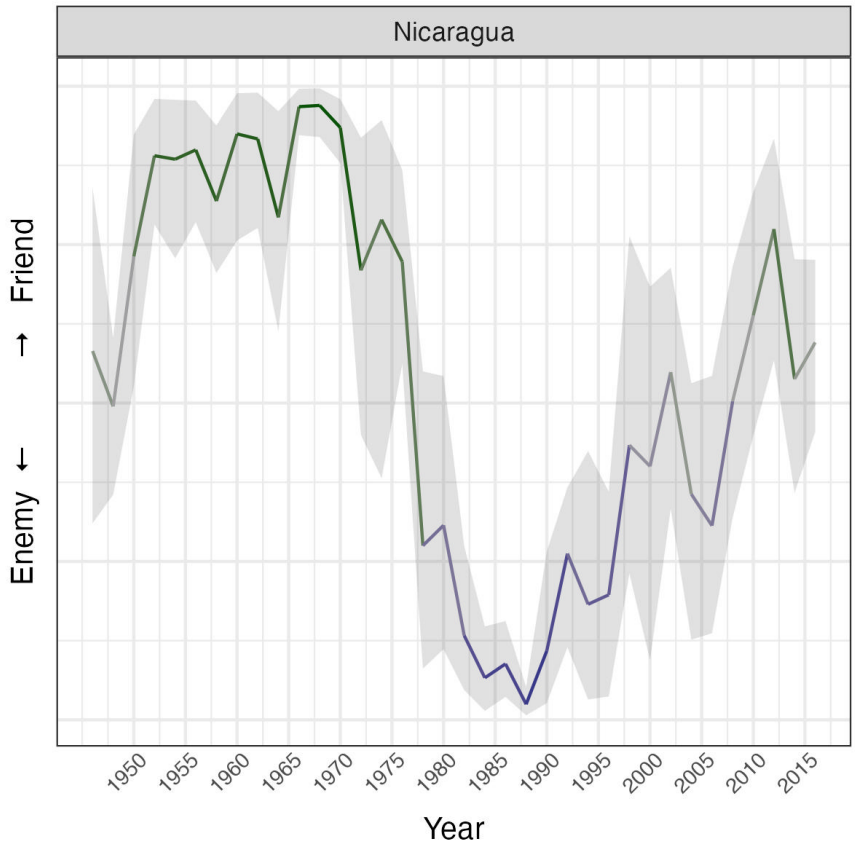
enemy	**them**	**france**	**russia**	**democrat**	**immigration**
enemies	him	frances	russian	democratic	immigrants
foe	these	germany	russians	democrats	immigrant
hostile	they	belgium	russias	republican	naturalization
bombardment	us	italy	soviet	republicans	migration
aggressor	their	switzerland	china	undemocratic	aliens
aggression	those	austria	soviets	party	deportation
attack	themselves	britain	prussia	politician	emigration
counterattack	yourselves	spain	czechoslovakia	candidate	migratory
aggressors	you	united_kingdom	germany	democracy	conflagration
conquer	whom	australia	japan	republicanism	migrate

4.5.2 Friendship and Collective Identity Projections in Bilateral Relations

The projections for friendship and collective identity capture important shifts in bilateral relations of the United States with other countries. This section offers two vignettes to illustrate these dynamics. Figure 4.8 plots the friendship projection of Nicaragua from 1945 through 2017. Following World War II until the early 1970s, the United States and Nicaragua enjoyed generally warm relations, with the U.S. supporting the Somoza government despite concerns about leftist movements in Latin America. Initially, the U.S. provided military assistance to the Somoza regime, but it eventually ceased due to human rights concerns. Nicaragua is projected close to the "friend" pole during this period. In the 1970s, the rise

of the Sandinista movement, which eventually overthrew the Somoza government in 1979, coincides with a significant decrease in the country's friendship projection.

Figure 4.8: Friendship Projection of Nicaragua



The 1980s are marked by a strong "enemy" projection for Nicaragua, aligning with the period in which the Sandinista government was in power. The Sandinista government pursued socialist policies and formed strong ties with the Soviet Union and Cuba, and in 1986, President Ronald Reagan even claimed that the "true enemy" of the Sandinistas was the United States.[44] The 1990s witnessed warmer relations between the United States and Nicaragua as the country transitioned to democracy and received substantial aid from the United States, with President Bill Clinton describing them as "neighbors and friends" in 1999.[45] This shift

[44]Reagan 1986.
[45]Gerstenzang 1999.

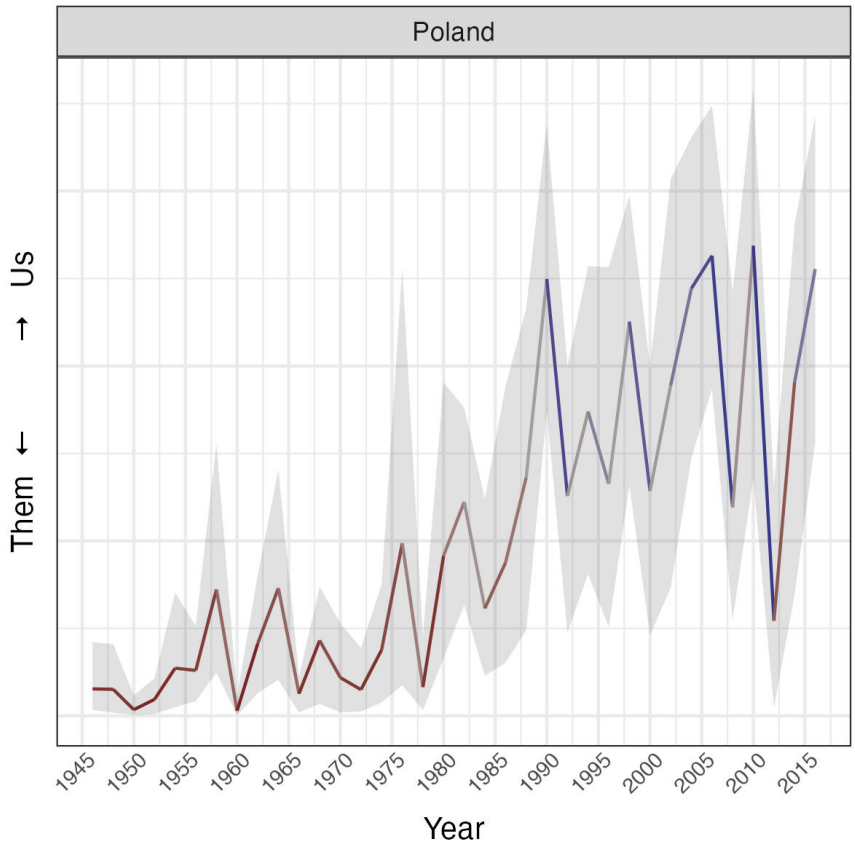
is reflected in Nicaragua's friendship projection moving towards the friend pole. Throughout the 2000s and 2010s, the U.S. maintained moderate relations with Nicaragua, characterized by a strong economic relationship, despite more recent concerns over democratic backsliding and the concentration of power under the Ortega government.

This methodological approach provides insights that would be difficult to obtain through alternative methods. Dictionary-based methods would likely be incapable of producing robust results. For instance, in the case of Nicaragua, there is a median of only one sentence per session of Congress that contains both the words "Nicaragua" and "friend". Using alliance status as a formal proxy would also yield uninformative outcomes, as there was no change in alliance status between Nicaragua and the United States from 1947 onwards.[46] While examining voting behavior at the United Nations could serve as a potential alternative approach, it primarily measures shared foreign policy preferences and does not directly assess perceptions of friendship or identity between any two states.

Figure 4.9 displays the collective identity projection for Poland from 1945 onwards. Poland exhibits a strong "them" projection from 1945 through approximately 1980, coinciding with its inclusion in the Soviet Union up until the emergence of the Solidarity Movement. This movement, which was heavily supported by the United States, aimed to dissolve authoritarian rule in Poland and ultimately helped facilitate Polish independence in 1989. Throughout the 1980s, Poland's collective identity projection moves towards the "us" pole. In 2002, President George W. Bush noted that Poland and the United States were "joined by a commitment to helping each other along freedom's road," and would together "complete the unification of Europe."[47] It is following this period and Poland's accession to the European Union in 2004 that the country's collective identity projection reaches its highest points in the mid-2000s.

[46] Gibler 2009.
[47] Bush 2002.

Figure 4.9: Collective Identity Projection of Poland



4.6 Friendship and Collective Identity Between States

4.6.1 Friendship

What explains variation in perceptions of friendship over time? Below, I present a variety of OLS models to examine the correlates of friendship. It is important to note that the models in the following analysis do not aim to establish causation, but rather to explore the characteristics of states that are perceived as friends, enemies, or have a shared identity with the United States.

The dependent variable in these models is the percentile rank of a state's friendship projection. These ranks are based on pooling the projections across the entire data set, such

that the lowest projection of any country in any year is scored as 0 and the highest projection of any country in any year is scored as 100. I include up to seven covariates in the models. The first covariate is *MID (Last 20 Years)*, which is a dummy variable indicating if the U.S. engaged in a militarized interstate dispute (MID) with the state in the preceding 20 years.[48] Previous studies have demonstrated negative correlations between MIDs and measures of geopolitical affinity, and the inclusion of this variable in the models offers insights into the relationship of military conflict with friendship and collective identity.[49]

Next, considering that states trade with both their friends and enemies, and recognizing that shared membership in international institutions can influence collective identity formation, I introduce two new variables. *Total Trade Flows (log)* takes the natural log of the volume of bilateral trade flows between the state and the U.S.[50] The other variable is *Alliance*, a dummy variable indicating if there is any type of alliance between U.S. and the other state that year, including defense, neutrality, nonaggression, and entente treaties.[51]

Common Language is a dummy variable indicating if there is a shared ethnic language between the countries. Prior research has identified language as an important factor in the formation of both friendship and collective identity.[52] *Distance (log)* takes the natural log of the distance in kilometers between the largest city by population of the two countries.[53] This variable reflects the changing importance of various regions throughout the 20th century, for reasons including shifts in the United States' grand strategy. *POLITY (Difference)* is the difference in POLITY2 scores of the state and the United States. Regime type has been linked to geopolitical affinity measures in previous studies.[54] Lastly, *CINC (scaled)* is the Composite Index of National Capability (CINC) score of the foreign state, which is a

[48]Palmer et al. 2022.
[49]M. A. Bailey, Strezhnev, and Voeten 2017; Gartzke 2000.
[50]I add 1 before taking the natural log.
[51]Gibler 2009.
[52]Durkin and Conti-Ramsden 2007; Edwards 2009; Heller 2003.
[53]Maddalena Conte, Cotterlaz, and Mayer 2022.
[54]M. A. Bailey, Strezhnev, and Voeten 2017; Marshall 2020.

measurement of the state's material capabilities. Given the pivotal role material capability often plays in the study of international politics, it is plausible that it may also have relevance in understanding the formation of friendship, enmity, or identity.

Table 4.2 presents the results of the first set of OLS models. In Model 1, *MID (Last 20 Years)* exhibits a coefficient of -12.783, indicating that having a recent militarized dispute correlates with approximately a 12.8 point percentile rank drop in friendship, all else equal. When adding year fixed effects in Model 2 and country and year fixed effects in Model 3, *MID (Last 20 Years)* maintains negative coefficients and remains statistically significant at conventional levels. However, the coefficient size decreases by about half to -7.011 in Model 3. These findings confirm the unsurprising result that militarized disputes are negatively associated with friendship. The *POLITY Difference* variable also has a negative effect across all models. In Model 3, after controlling for time and country-invariant factors, a difference of about 5 POLITY points corresponds to a 1 percentile rank decrease in friendship. Specifically, a difference of 20 POLITY points, the maximum in the data set, yields a change of approximately 3.9 percentile ranks.

Table 4.2: Correlates of Friendship Projections

	Friendship Percentile Rank		
	(1)	(2)	(3)
MID (Last 20 Years)	-12.783^{***}	-14.437^{***}	-7.011^{***}
	(1.049)	(0.967)	(0.960)
Total Trade Flows (Log)	-0.001	0.452^{***}	1.008^{***}
	(0.115)	(0.121)	(0.130)
Alliance	5.040^{***}	-1.101	3.652^{***}
	(0.931)	(0.929)	(1.266)
Common Language	-0.248	0.220	
	(0.813)	(0.747)	
Distance (Log)	-5.199^{***}	-7.555^{***}	
	(0.808)	(0.768)	
POLITY Difference	-0.486^{***}	-0.629^{***}	-0.194^{***}
	(0.057)	(0.054)	(0.067)
CINC (Scaled)	-5.381^{***}	-4.652^{***}	-1.457^{*}
	(0.461)	(0.448)	(0.825)
Constant	102.025^{***}	117.021^{***}	28.880^{***}
	(7.619)	(8.617)	(6.072)
Year Fixed Effects	No	Yes	Yes
Country Fixed Effects	No	No	Yes
Observations	4,159	4,159	4,160
R^2	0.149	0.300	0.562
Note:			*p<0.1; **p<0.05; ***p<0.01

The coefficients on *Total Trade Flows (log)* are positive in all models and reach conventional levels of statistical significance when adding fixed effects in Models 2 and 3. Model 3 is particularly important since it accounts for changes in the composition of countries in the data set over time. The coefficient of 1.008 indicates that a 1% increase in total trade

volume is associated with about a 0.01 increase in friendship percentile rank, holding all else constant. *Alliance* also has a positive coefficient in Model 3, showing that any shared alliance corresponds to a 3.7 percentile rank increase in friendship.

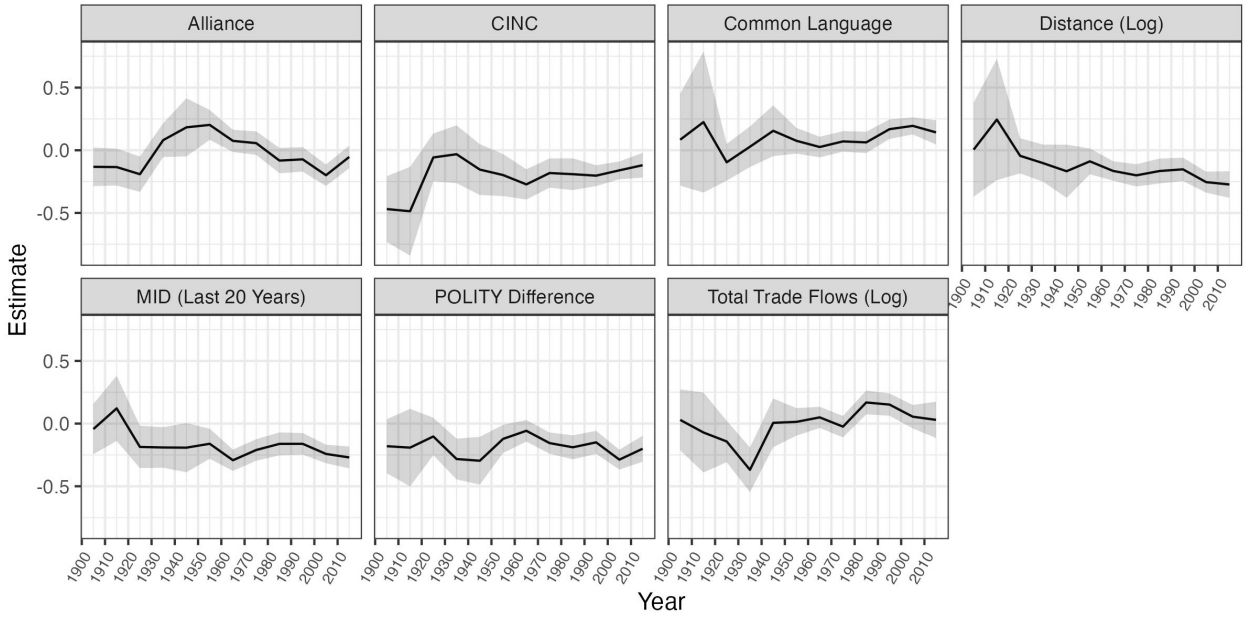
Distance (log) exhibits a negative effect in Models 1 and 2, where a 1% increase in the geographic distance between states corresponds with a drop in friendship of .0520 and .0756 percentile ranks, respectively. This result indicates that geographic neighbors are perceived more as friends than distant states, holding all else constant. Finally, *CINC (Scaled)* has a negative coefficient in all models, indicating that the U.S. is less likely to perceive materially stronger states as friends.

Figure 4.10 expands on this analysis by examining the relationship between the covariates and friendship over time. I separate the data by decade and estimate an OLS model for each decade worth of data per the specification of Model 1. I also standardized all variables before estimating the models to better assess the relative importance of the covariates. Specifically, a coefficient of 1 indicates that a standard deviation increase in the covariate is associated with a standard deviation increase in friendship.

The effect of *Alliance* has changed substantially over time. Between the interwar period and the 1970s, states with whom the U.S. had a shared alliance experienced higher friendship projections, all else equal. However, this trend reversed in the 1980s and has since remained negative, reaching an all-time low in the 2000s. This coincided with a period of significant tension between the United States and its European allies over the Iraq War, with U.S. Defense Secretary Donald Rumsfeld referring to France and Germany as "problems" due to their lack of support for the U.S. invasion of Iraq.[55] The effects of all other variables have remained relatively stable since World War II. *CINC*, *Distance (Log)*, *MID (Last 20 Years)*, and *POLITY Difference* have exhibited negative coefficients across all decades since the 1920s, while *Total Trade Flows (Log)* has remained positive from the 1940s onward. While

[55] *Secretary Rumsfeld Briefs at the Foreign Press Center* 2003.

Figure 4.10: Standardized Regression Coefficients for Friendship Model



the estimates from the earlier decades of coverage show greater variance, this can likely be attributed to the significantly smaller sample size during that time. Overall, these findings suggest that the factors associated with friendship have remained relatively constant over time.

4.6.2 Collective Identity

Next, I investigate the correlates of perceptions of collective identity. Table 4.3 presents the results of OLS models which take the same covariates as the models presented above in Table 4.2, but with the percentile rank of collective identity as the dependent variable.

Across all models, *MID (Last 20 Years)* shows a positive coefficient, but its estimates do not reach conventional levels of statistical significance in Models 5 and 6. *Total Trade Flows (Log)* has a positive effect in all model specifications, albeit with a smaller effect size in Model 6 compared to the effect on friendship in Model 3. *Common Language* also exhibits positive coefficients in Models 4 and 5, indicating that a shared language corresponds to a

Table 4.3: Correlates of Collective Identity Projections

	Collective Identity Percentile Rank		
	(4)	(5)	(6)
MID (Last 20 Years)	1.842*	0.846	0.022
	(1.059)	(0.979)	(1.012)
Total Trade Flows (Log)	1.740***	0.284**	0.616***
	(0.116)	(0.122)	(0.137)
Alliance	−0.474	3.546***	2.298*
	(0.940)	(0.940)	(1.335)
Common Language	2.034**	3.187***	
	(0.821)	(0.756)	
Distance (Log)	4.628***	2.457***	
	(0.816)	(0.777)	
POLITY Difference	−0.093	0.029	0.004
	(0.057)	(0.055)	(0.071)
CINC (Scaled)	−1.253***	0.747	−0.430
	(0.466)	(0.454)	(0.870)
Constant	−24.668***	6.394	26.800***
	(7.694)	(8.721)	(6.403)
Year Fixed Effects	No	Yes	Yes
Country Fixed Effects	No	No	Yes
Observations	4,159	4,159	4,160
R^2	0.071	0.232	0.479
Note:			*p<0.1; **p<0.05; ***p<0.01

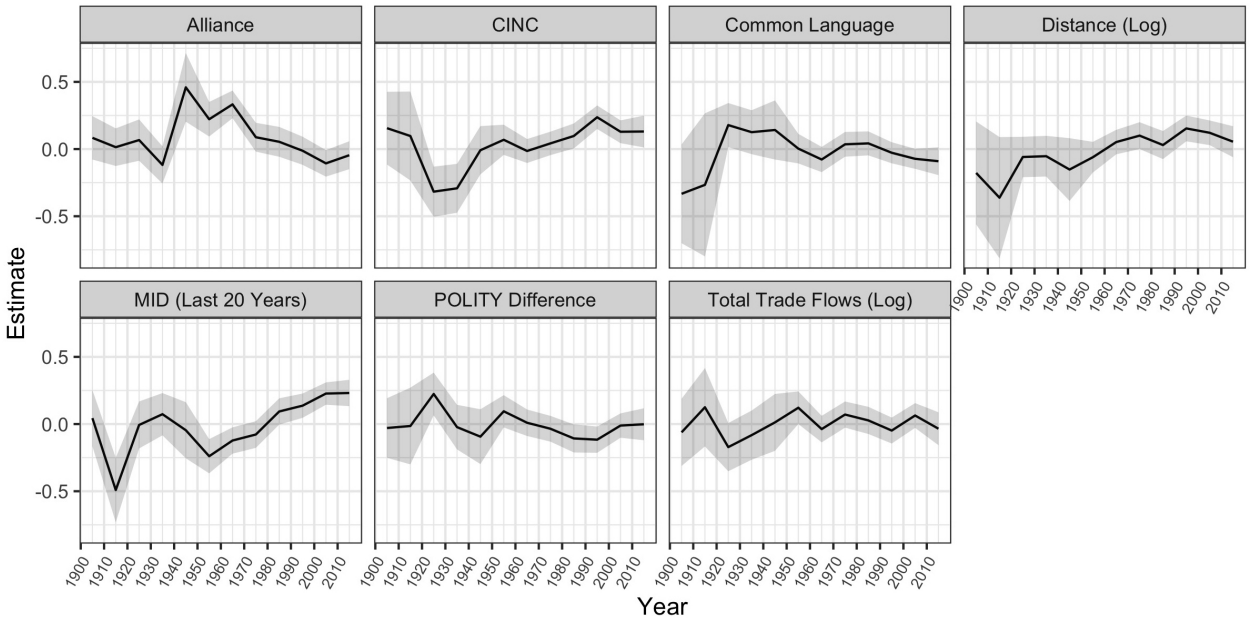
higher collective identity. The positive coefficient on *Distance (Log)* is perhaps a surprising result, indicating that states which are geographically distant are perceived to have a stronger collective identity with the United States, all else being equal. The coefficient of 2.457 in Model 5 indicates that a 1% increase in distance is associated with a 0.025 percentile rank increase in collective identity. Note that this is the opposite direction of the effect of *Distance (Log)* in the friendship models, suggesting that neighboring countries may be perceived as friends with the U.S. but not share a collective identity, while more distant countries may be perceived to have less friendship but more collective identity.

POLITY Difference is statistically indistinguishable from zero in all models. While regime type is a strong indicator of friendship, it appears to have little bearing on collective identity in the fully specified model. Lastly, whereas the effect of *CINC (Scaled)* was negative in all friendship models, it varies significantly across the model specifications for collective identity.

Figure 4.11 plots the effects of these covariates over time for standardized OLS models without fixed effects. The magnitude and direction of these effects vary substantially more over time for collective identity they do in the friendship models. The effect of *Alliance* has significantly decreased since the 1940s and is now statistically indistinguishable from zero. This finding suggests that there was a collective identity between the U.S. and its alliance partners in the post-war period that no longer exists today.

In contrast, the coefficients for *CINC* and *Distance (Log)* have generally increased over the data period, with the directions of their effects changing from negative to positive in the mid-20th century. These results indicate that while the U.S. had a stronger sense of collective identity with materially weak, geographically proximate states at the beginning of the 20th century, there is now a diminished sense of collective identity with these states. *Common Language* was primarily positive through the interwar period, but has since become statistically indistinguishable from zero. The effect of *MID (Last 20 Years)* has oscillated

Figure 4.11: Standardized Regression Coefficients for Collective Identity Model



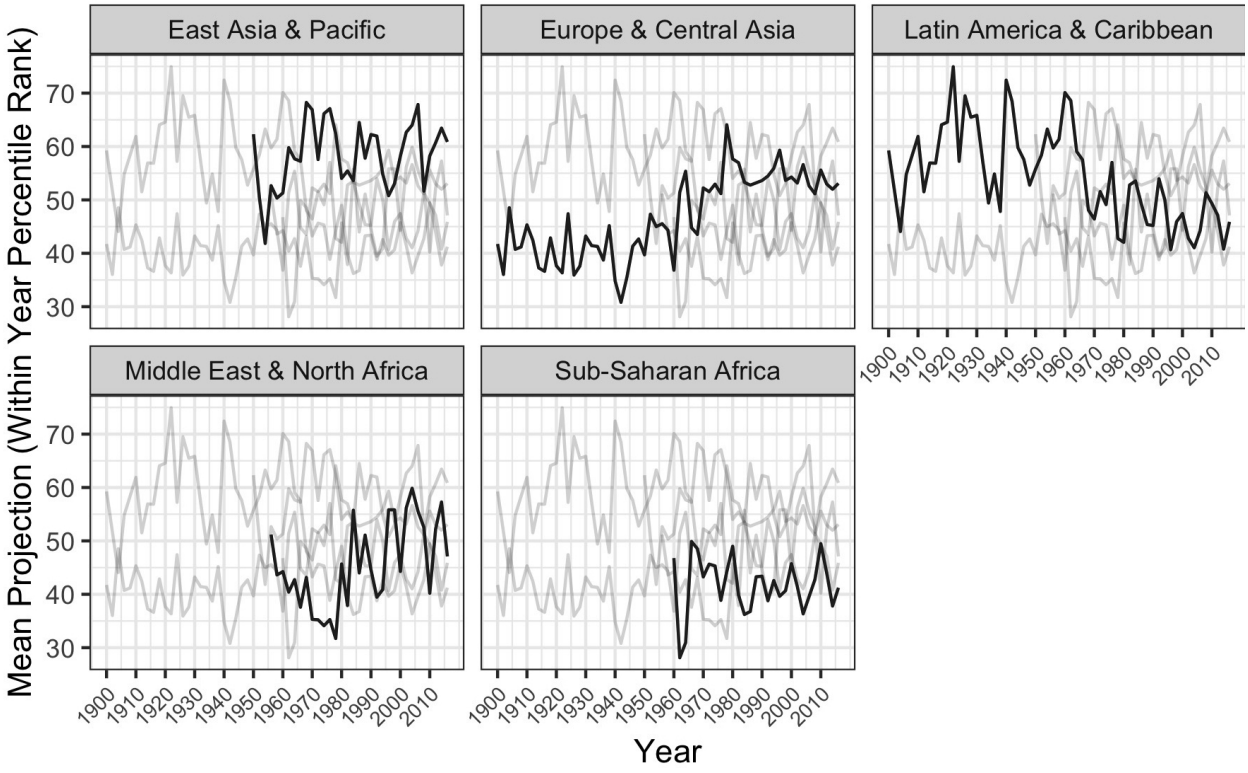
substantially over time. While it exhibited a negative effect on collective identity in the post-war period, its direction reversed in the 1980s and has been positive ever since. *POLITY Difference* and *Total Trade Flows (log)* have not reached conventional levels of statistical significance in most decades.

Overall, these results indicate that material considerations and conflict are more strongly associated with perceptions of collective identity than perceptions of friendship. In the post-war period, shared alliances and an absence of recent militarized disputes were the strongest correlates of collective identity. More recently, a history of militarized disputes is associated with *increased* collective identity, and materially strong states are perceived as more like "us".

Figure 4.12 displays the mean collective identity projection per region, measured by within-year percentile ranks. The results indicate that European and Central Asian countries experienced a significant growth in collective identity from approximately 1940 through 1975. This trend supports the hypothesis that a "North Atlantic" collective identity emerged

between the United States and European countries during and after World War II.[56] In contrast, Latin American and Caribbean countries have shown a decreasing trend in collective identity since the 1960s.

Figure 4.12: Mean Collective Identity Projection Per Region (Minimum 10 Countries)



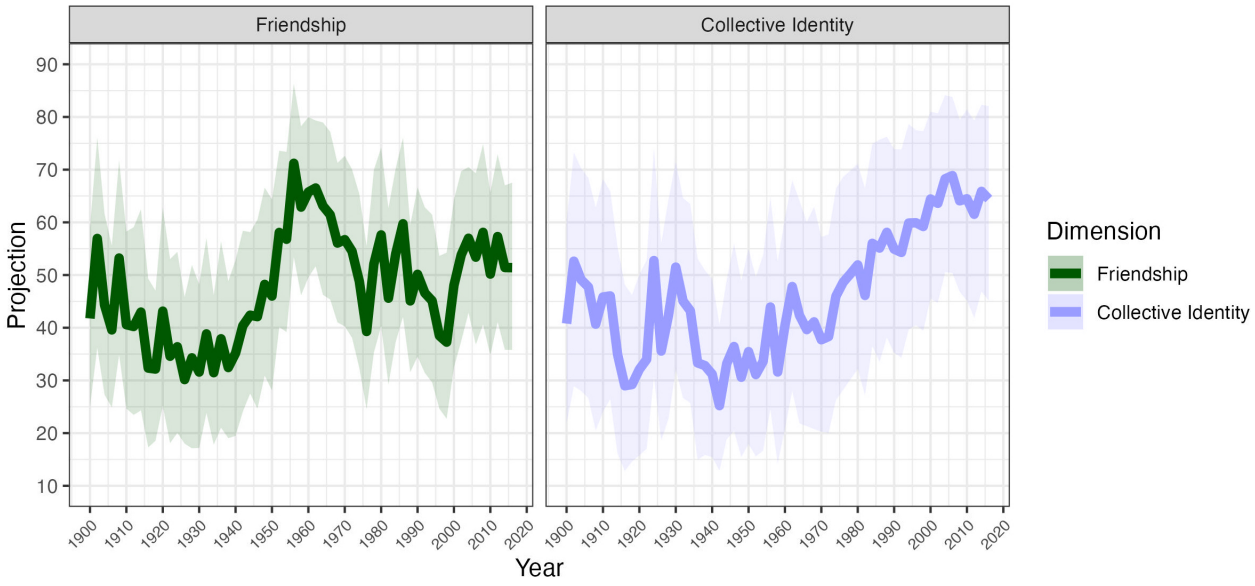
4.6.3 Friendship and Collective Identity

The third section of this analysis focuses on the relationship between friendship and collective identity. Figure 4.13 displays the mean pooled percentile rank of all countries for both dimensions.

The mean friendship projection of all countries experienced a significant increase during and after World War II, ultimately reaching its peak in the late 1950s. The period after

[56]Deutsch 1957; Hemmer and Katzenstein 2002.

Figure 4.13: Mean Annual Projection of All Countries

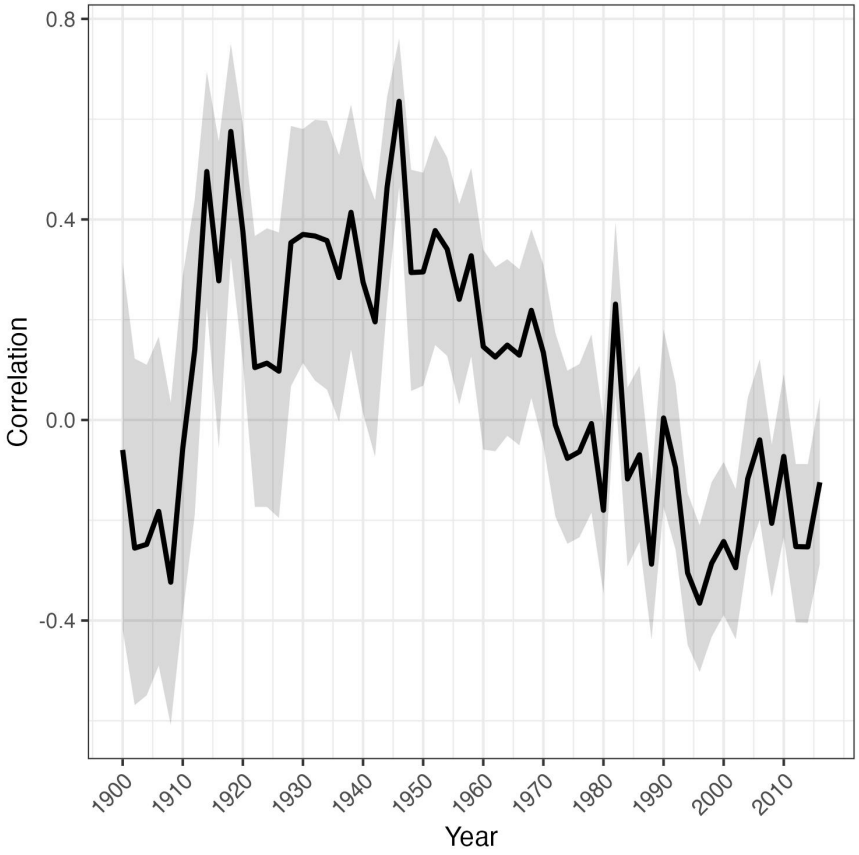


World War II until the Vietnam War saw the highest level of friendship, which has remained generally lower since then. These findings suggest that, on average, the United States perceived foreign states as particularly friendly during the postwar era. On the other hand, the mean collective identity projection of all states has consistently increased since the beginning of World War II, indicating that the United States has continuously perceived a stronger sense of collective identity with foreign nations over the past eight decades.

Figure 4.14 depicts the correlation between within-year percentile ranks of friendship and collective identity. Throughout much of the early to mid-20th century, a positive correlation existed between these projections. This implies that countries with whom the U.S. perceived a collective identity were more likely to be considered as friends. This observation mirrors the significant collaboration between the U.S. and its allies in the early Cold War era, highlighted by the establishment of various international institutions such as NATO.

However, following this period, the positive correlation began to decline and eventually disappeared entirely, resulting in a negative correlation between friendship and collective

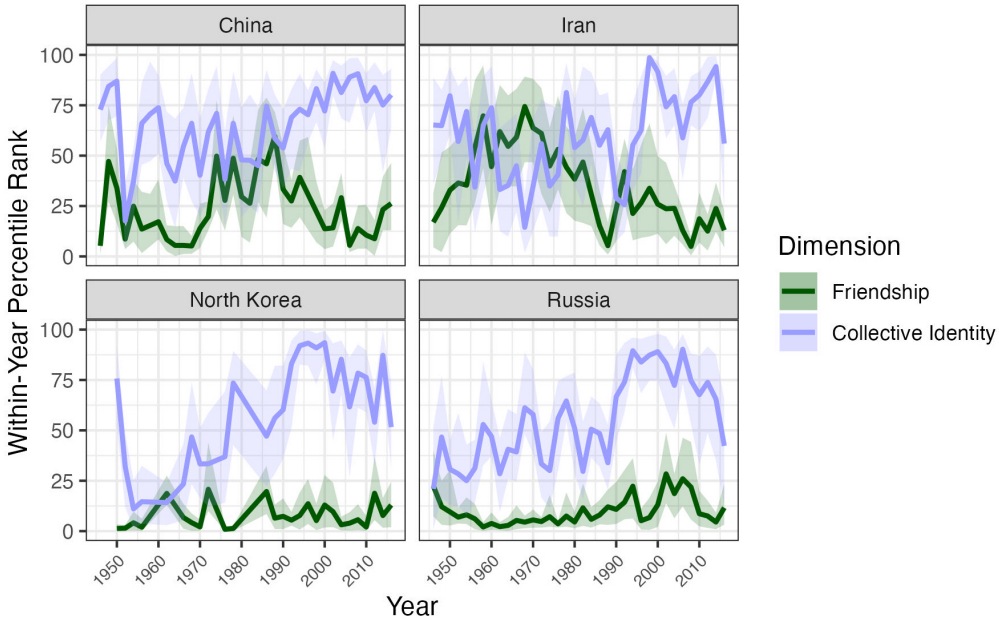
Figure 4.14: Correlation of Friendship and Collective Identity Projections



identity in recent years. This suggests two phenomena. First, the U.S. may have more "transactional" perceptions of friendships, where they perceive less of a collective identity with the other state. Second, the diminished positive correlation between friendship and collective identity suggests the emergence of a collective identity among adversaries. This trend is generally aligned with the emergence of détente during the late Cold War era, in which tensions with the Soviet Union and Eastern bloc countries gave way to more cooperation and diplomatic efforts. Furthermore, in more recent times, the U.S. has been more apt to openly critique its friends and allies. For instance, significant trade-related tensions arose with Japan in the 1980s, and key military partners were criticized for their lack of support of the War in Afghanistan. There has even been a noticeable rise in perceptions of collective

identity between the United States and some of its notable adversaries in recent decades (see Figure 4.15). In summary, while there was a convergence of friendship and collective identity in the early to mid-20th century, these concepts have become dissociated in more recent years.

Figure 4.15: Friendship and Collective Identity Projections of Notable Countries



4.6.4 Additional Sources of Variation in Country Projections

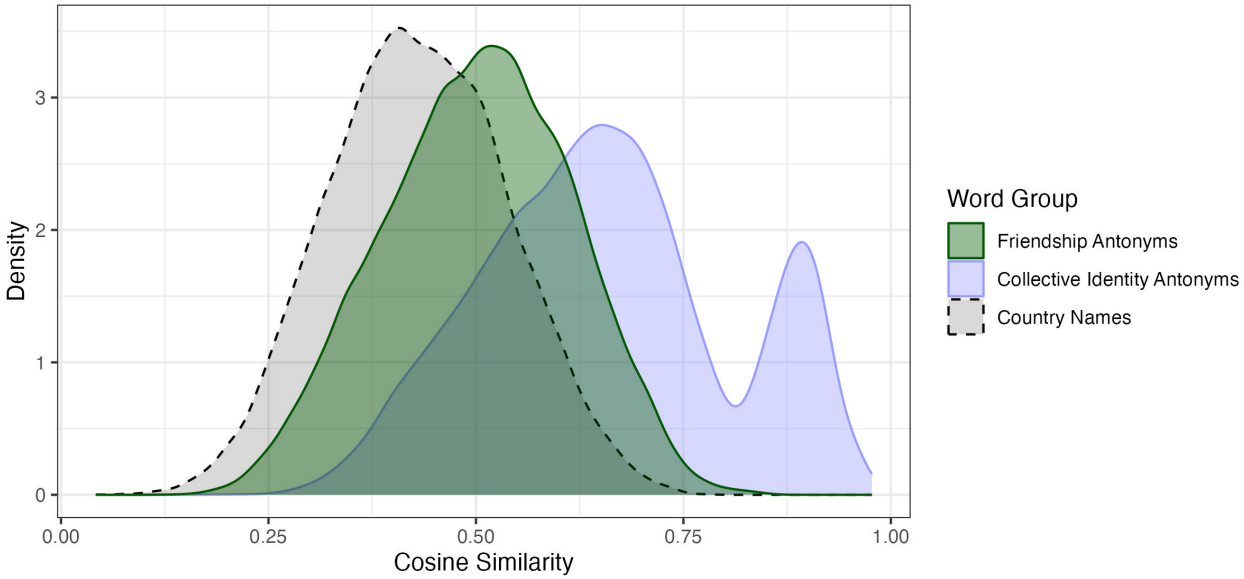
When analyzing the projections for friendship and collective identity, it is important to consider other potential sources of variation. One such source could be changes in the dimensional vector $\boldsymbol{d}$ on which I project the country word vectors $\boldsymbol{w}$, which would produce variation in the projections between time periods. In the preceding analysis, I assume that the dimensional vectors do not change, meaning that the usage of the antonym pairs relative to each other is constant over time. This allows for the comparison of country projections across different time periods.

To test the validity of this assumption, I measure changes in the word vectors of the antonym pairs between aligned embedding spaces. Embedding alignment permits the direct comparison of word vectors between embedding spaces by mapping embedding spaces to each another.[57] While the relative positions of word vectors can be compared across embedding spaces, the word vectors themselves are incomparable without alignment. To align the embedding spaces, I use a set of stop words with meanings that I assume remain constant over time (see Appendix B.6). After this procedure, changes in the word vectors between the aligned spaces indicate changes in word meaning. I calculate the cosine similarity of each antonym word and each country word between all time periods and compare the resulting distributions in Figure 4.16.

The antonym pairs related to collective identity have the highest mean cosine similarity between time periods, with a value of 0.652. This indicates that the meanings of these words change the least over time. The friendship antonym words exhibit slightly less stability, with a mean cosine similarity of 0.502. The country words have the lowest cosine similarity value at 0.430, indicating that their meanings have changed the most over time. These findings suggest that changes in the meaning of country words, rather than changes in the meaning

[57] Kalinowski and An 2020; Milbauer, Mathew, and J. Evans 2021.

Figure 4.16: Distribution of Word Vector Cosine Similarities Between Aligned Models

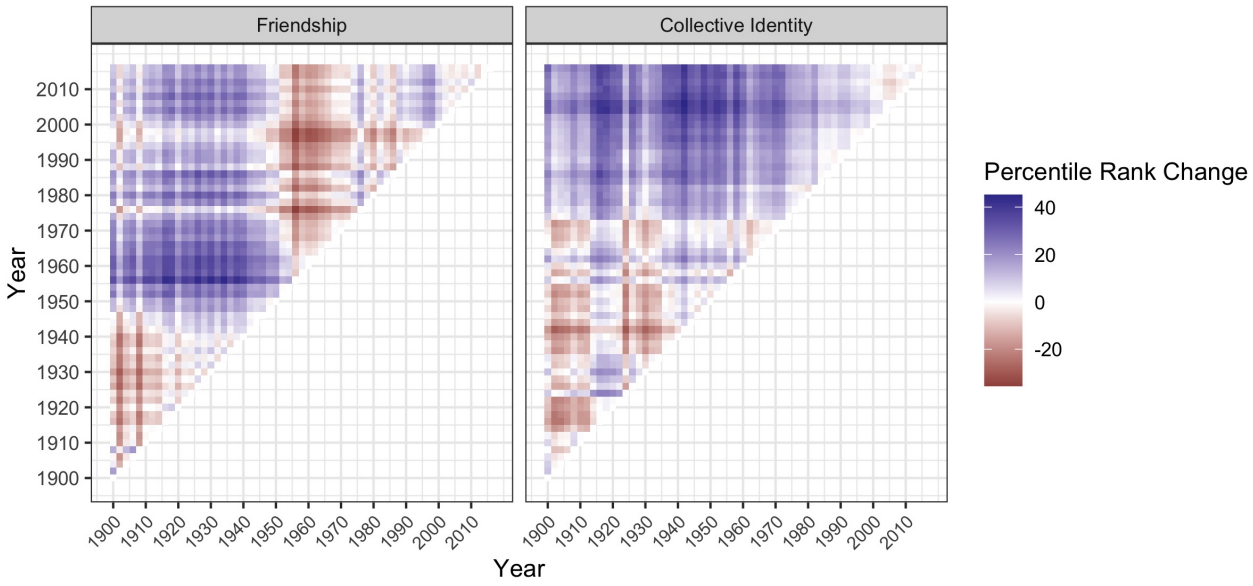


of antonym pairs used to construct the dimensional vectors, produce at least the majority of the variation observed in the data.

Another factor that could contribute to the variation observed in this analysis is change in the composition of countries included in the data set. As new countries come into existence and old ones disappear, we might observe aggregate trends in the projections even if the projections of individual countries did not change over time. Figure 4.17 plots within-country change in the friendship and collective identity projections. Each square depicts the mean change in pooled percentile rank from *Year X* to *Year Y* for all countries which are included in both years in the data set. Beginning with the left panel, we observe an upwards shift in friendship ranks beginning in 1942 for each cohort of countries present in the data set before that year. This indicates that countries which were already in the data set before 1942 experienced an upward shift in friendship projections, providing evidence that the overall trend observed in the data was not solely due to the inclusion of new countries. We observe a similar result in the right panel with the collective identity projection data, suggesting

that countries which existed in the early 20th century also experienced an upward trend in collective identity starting later in the century.

Figure 4.17: Mean Change in Projections by Year



4.7 Conclusion

The concepts of friendship and collective identity commonly appear in the context of international politics, but measuring them in empirical studies has proven to be a challenging task. This chapter represents a first step in quantifying perceptions of friendship and collective identity between states. Employing the dimensional approach to word embeddings, I develop discursive measures of friendship and collective identity, analyzing their dynamics between the United States and 192 other countries over a span of 117 years. The findings reveal that perceptions of friendship and collective identity have exhibited distinct evolutionary patterns over time. Notably, many states that are considered enemies of the United States are also associated with their in-group in recent years, indicating the presence of a shared identity among adversaries in the realm of international politics.

There are several limitations to the quantitative approach I present in this chapter. First, I assume that perceptions of friendship and collective identity can be inferred solely through the speech of legislators. However, friendship and identity are multifaceted concepts, rooted in various practices between social actors, where speech is only one aspect of how these concepts are expressed and experienced.[58] By relying on a text-based approach, this study focuses on a single component of the complex interactions that contribute to the formation of friendship or shared identity. Additionally, by focusing on the speech of legislators, I neglect the perspectives of other important actors, such as heads of state or the general public, who may have different views on friendship or identity between states. To address this limitation, a potential avenue for future research could involve using corpora related to other types of actors to gain a more comprehensive understanding of how friendship and identity are perceived and practiced.

Furthermore, expanding the computational scope of this research could lead to more robust findings. Allocating greater resources to optimize the hyperparameter choices when constructing the embedding models might result in more meaningful and stable embeddings and more reliable measures of friendship and collective identity. Following the recommendations of Rodriguez and Spirling (2022), researchers could experiment with different hyperparameter choices to ensure stable results before proceeding with the analysis. Currently, this study presents an analysis based on a single set of hyperparameters for the embedding models.

This study makes three primary contributions to international relations research. First, I challenge the notion that "othering" another party is synonymous with making them an enemy. My findings suggest that although these concepts may have some correlation, their relationship has changed over time. In recent years, I find a negative relationship between perceptions of friendship and collective identity, indicating that states may have enemies

[58] Berger and Luckmann 1990; Bourdieu 1990; Gouvard, Goldberg, and S. B. Srivastava 2023.

with whom they share a common identity. For theories of international politics which take social identity as an essential component in explaining state behavior, my findings underscore the significance of avoiding the language of friendship as a rhetorical device when analyzing state identity. This chapter also speaks to Wendt (1999)'s theory of alternative structures of anarchy, which revolves around the dominant identity of states in the international system.[59] My findings suggest that it may be challenging to identify a single "dominant" identity of states, as states may collectively identify with both strong enemies and strong friends at a given time.

This research also has implications for future studies that wish to incorporate measures of geopolitical affinity into quantitative models. To ensure clarity and consistency, scholars should provide explicit definitions of "geopolitical affinity" within their research. When researchers are primarily concerned with a state's foreign policy preferences in isolation, established measures like UN Ideal Point likely remain appropriate. However, for those interested in considering characteristics of dyadic relationships, adopting a geopolitical affinity measure that is more in line with the friendship or collective identity measures presented in this chapter would be more suitable. The field would greatly benefit from a more comprehensive dyadic data set of friendship and collective identity beyond the U.S. focus of this study. On this point, one avenue for future research may be to leverage a corpus of UNGA speeches and a pretrained embedding model to generate dyadic projections between a wider set of states.

Finally, the dimensional approach to word embeddings is widely transferable and holds potential for exploring many concepts in IR. For instance, a similar study could be conducted to measure perceptions of power between states, aligning more closely with constructivist theories that emphasize actor perception. In the realm of international political economy, scholars could investigate how perceptions of affluence or responsibility relate to the forma-

[59] Wendt 1999.

tion of trade relationships and interactions within international institutions. This method could also be applied to the study of political leaders in IR, for instance, by measuring how the age of leaders correlates with perceptions of their age among their constituents.[60]

Alexander Wendt remarked, "if scholars are willing to treat states as enemies, then it makes no sense to apply a different standard to 'friend.'"[61] Collective identity has been rigorously theorized in the field of IR; friendship, however, has received less scholarly attention. Perhaps the most significant contribution of this chapter is taking the concept of friendship seriously as a subject of study in international relations. While enmity is likely of greater consequence to the dynamics of international politics, friendship undoubtedly plays a central role in many of these interactions as well.

[60]Byun and Carson 2022.
[61]Wendt 1999, p. 298.

Bibliography

A Big Data Approach to Assessing the Impact of Social Norms: Reporting One's Exercise to a Social Media Audience - Christopher J. Carpenter, Chandra S. Amaravadi, 2019 (2021). URL: https://journals.sagepub.com/doi/full/10.1177/0093650216657776?casa_token=D1dQdHxLd_UAAAAA%3A3KhjnE7jcrHz7DVVvZeiElps_3iKu5fTg1gESZ1F23h5dDP_e49-Sazi3xp74ax-FPz-d7XZe3c (visited on 10/13/2021).

About | 2Africacable.Com (2021). 2africacable.com Fin. URL: https://www.2africacable.com/about (visited on 10/12/2021).

Acemoglu, Daron et al. (Sept. 2019). *Too Much Data: Prices and Inefficiencies in Data Markets*. w26296. Cambridge, MA: National Bureau of Economic Research, w26296. DOI: 10.3386/w26296. URL: http://www.nber.org/papers/w26296.pdf (visited on 02/16/2021).

Adler, Emanuel and Michael Barnett (Oct. 28, 1998). "Security Communities in Theoretical Perspective". In: *Security Communities*. Ed. by Emanuel Adler and Michael Barnett. 1st ed. Cambridge University Press, pp. 3–28. ISBN: 978-0-521-63051-1 978-0-521-63953-8 978-0-511-59866-1. DOI: 10.1017/CBO9780511598661.001. URL: https://www.cambridge.org/core/product/identifier/CBO9780511598661A009/type/book_part (visited on 03/29/2023).

Ahmad, Najam and Kevin Salvadori (May 14, 2020). *Building a Transformative Subsea Cable to Better Connect Africa*. Facebook Engineering. URL: https://engineering.fb.com/2020/05/13/connectivity/2africa/ (visited on 10/12/2021).

Airbnb (Mar. 11, 2022). *An Update on Our Response to the Crisis in Ukraine*. URL: https://news.airbnb.com/an-update-on-our-response-to-the-crisis-in-ukraine/#:~:text=to%20the%20guests.-,Airbnb%20continues%20to%20temporarily%20waive%20our%20Host%20and%20guest%20service,c)(3)%20nonprofit%20organization..

Alexander, Dan and Bryan Rooney (Mar. 1, 2019). "Vote-Buying by the United States in the United Nations". In: *International Studies Quarterly* 63.1, pp. 168–176. ISSN: 0020-8833, 1468-2478. DOI: 10.1093/isq/sqy059. URL: https://academic.oup.com/isq/article/63/1/168/5303354 (visited on 03/29/2023).

Allam, Ahmed, Peter Johannes Schulz, and Kent Nakamoto (2014). "The Impact of Search Engine Selection and Sorting Criteria on Vaccination Beliefs and Attitudes: Two Experiments Manipulating Google Output". In: *Journal of medical internet research* 16.4, e2642.

Alt, James E. et al. (1999). "Asset Specificity and the Political Behavior of Firms: Lobbying for Subsidies in Norway". In: *International Organization* 53.1, pp. 99–116. ISSN: 0020-

8183, 1531-5088. DOI: 10.1162/002081899550823. URL: https://www.cambridge.org/core/product/identifier/S0020818399440640/type/journal_article (visited on 08/10/2023).

Amer, Mildred (1993). *The Congressional Record; Content, History and 'Issues*. Congressional Research Service. URL: https://www.llsdc.org/assets/sourcebook/crs-93-60.pdf.

"America's Friends and Enemies" (Feb. 2, 2017). In: *YouGovAmerica*. URL: https://today.yougov.com/topics/international/articles-reports/2017/02/02/americas-friends-and-enemies.

Anderson, Benedict R. O'G (2016). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Revised edition. London New York: Verso. 240 pp. ISBN: 978-1-78478-675-5 978-1-84467-484-8 978-1-78168-359-0.

Angell, Robert Cooley (1969). *Peace on the March; Transnational Participation*. New York: Van Nostrand Reinhold Co. ISBN: 978-0-442-00333-3.

Aristotle and Terence Irwin (1999). *Nicomachean Ethics*. 2nd ed. Indianapolis, Ind: Hackett Pub. Co. 360 pp. ISBN: 978-0-87220-465-2 978-0-87220-464-5.

Arsenault, Amelia C. and Sarah E. Kreps (Feb. 14, 2022). "AI and International Politics". In: *The Oxford Handbook of AI Governance*. Ed. by Justin B. Bullock et al. 1st ed. Oxford University Press. ISBN: 978-0-19-757932-9 978-0-19-757935-0. DOI: 10.1093/oxfordhb/9780197579329.013.49. URL: https://academic.oup.com/edited-volume/41989/chapter/355439874 (visited on 08/09/2023).

Ash, Elliott, Daniel Chen, and Arianna Ornaghi (2021). "Gender Attitudes in the Judiciary : Evidence from U.S. Circuit Courts". In: *QAPEC Discussion Papers* 08.

Asher-Schapiro, Avi and Fabio Teixeira (Oct. 5, 2021). "Facebook down: What the Outage Meant for the Developing World". In: *Thomson Reuters Foundation*. URL: https://news.trust.org/item/20211005204816-qzjft/ (visited on 03/16/2023).

Bail, Christopher A. (2021). *Breaking the Social Media Prism: How to Make Our Platforms Less Polarizing*. 1st ed. Princeton: Princeton University Press. ISBN: 978-0-691-20342-3.

Bailey, Michael, Rachel Cao, et al. (Aug. 1, 2018). "Social Connectedness: Measurement, Determinants, and Effects". In: *Journal of Economic Perspectives* 32.3, pp. 259–280. ISSN: 0895-3309. DOI: 10.1257/jep.32.3.259. URL: https://pubs.aeaweb.org/doi/10.1257/jep.32.3.259 (visited on 06/28/2023).

Bailey, Michael, Patrick Farrell, et al. (July 2020). "Social Connectedness in Urban Areas". In: *Journal of Urban Economics* 118, p. 103264. ISSN: 00941190. DOI: 10.1016/

j . jue . 2020 . 103264. URL: https : / / linkinghub . elsevier . com / retrieve / pii / S0094119020300358 (visited on 06/28/2023).

Bailey, Michael, Abhinav Gupta, et al. (Mar. 2021). "International Trade and Social Connectedness". In: *Journal of International Economics* 129, p. 103418. ISSN: 00221996. DOI: 10 . 1016 / j . jinteco . 2020 . 103418. URL: https : / / linkinghub . elsevier . com / retrieve/pii/S0022199620301331 (visited on 06/28/2023).

Bailey, Michael A., Anton Strezhnev, and Erik Voeten (Feb. 1, 2017). "Estimating Dynamic State Preferences from United Nations Voting Data". In: *Journal of Conflict Resolution* 61.2, pp. 430–456. ISSN: 0022-0027. DOI: 10 . 1177 / 0022002715595700. URL: https : //doi.org/10.1177/0022002715595700 (visited on 12/21/2020).

Ballatore, Andrea, Mark Graham, and Shilad Sen (Sept. 3, 2017). "Digital Hegemonies: The Localness of Search Engine Results". In: *Annals of the American Association of Geographers* 107.5, pp. 1194–1215. ISSN: 2469-4452. DOI: 10 . 1080 / 24694452 . 2017 . 1308240. URL: https : / / doi . org / 10 . 1080 / 24694452 . 2017 . 1308240 (visited on 11/28/2021).

Bandy, Jack (Apr. 13, 2021). "Problematic Machine Behavior: A Systematic Literature Review of Algorithm Audits". In: *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW1), pp. 1–34. ISSN: 2573-0142. DOI: 10.1145/3449148. URL: https://dl.acm.org/doi/10.1145/3449148 (visited on 08/02/2023).

Baum, Matthew and Philip B. K. Potter (2015). *War and Democratic Constraint: How the Public Influences Foreign Policy*. Princeton, New Jersey: Princeton University Press. 258 pp. ISBN: 978-0-691-16498-4 978-0-691-16523-3.

Baum, Matthew A. and Philip B. K. Potter (Apr. 2019). "Media, Public Opinion, and Foreign Policy in the Age of Social Media". In: *The Journal of Politics* 81.2, pp. 747–756. ISSN: 0022-3816, 1468-2508. DOI: 10.1086/702233. URL: https://www.journals.uchicago.edu/doi/10.1086/702233 (visited on 07/31/2023).

— (June 1, 2008). "The Relationships Between Mass Media, Public Opinion, and Foreign Policy: Toward a Theoretical Synthesis". In: *Annual Review of Political Science* 11.1, pp. 39–65. ISSN: 1094-2939, 1545-1577. DOI: 10 . 1146 / annurev . polisci . 11 . 060406 . 214132. URL: https://www.annualreviews.org/doi/10.1146/annurev.polisci.11.060406.214132 (visited on 08/09/2023).

Bauman, Zygmunt et al. (June 1, 2014). "After Snowden: Rethinking the Impact of Surveillance". In: *International Political Sociology* 8.2, pp. 121–144. ISSN: 1749-5679. DOI: 10 . 1111/ips.12048. URL: https://doi.org/10.1111/ips.12048 (visited on 02/17/2021).

Bergemann, Dirk, Alessandro Bonatti, and Tan Gan (Jan. 31, 2021). "The Economics of Social Data". arXiv: 2004.03107 [cs, econ, q-fin]. URL: http://arxiv.org/abs/2004.03107 (visited on 02/16/2021).

Berger, Peter L. and Thomas Luckmann (1990). *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. New York: Anchor Books. 219 pp. ISBN: 978-0-385-05898-8.

Bessen, James (Aug. 1, 2020). "Industry Concentration and Information Technology". In: *The Journal of Law and Economics* 63.3, pp. 531–555. ISSN: 0022-2186, 1537-5285. DOI: 10.1086/708936. URL: https://www.journals.uchicago.edu/doi/10.1086/708936 (visited on 10/05/2021).

Bickel, Peter J. and David A. Freedman (Nov. 1, 1981). "Some Asymptotic Theory for the Bootstrap". In: *The Annals of Statistics* 9.6. ISSN: 0090-5364. DOI: 10.1214/aos/1176345637. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-9/issue-6/Some-Asymptotic-Theory-for-the-Bootstrap/10.1214/aos/1176345637.full (visited on 03/29/2023).

Bigo, Didier and Laurent Bonelli (2019). "Digital Data and the Transnational Intelligence Space". In: *Data Politics: Worlds, Subjects, Rights*. Routledge Studies in International Political Sociology. London ; New York: Routledge, Taylor & Francis Group. ISBN: 978-1-138-05325-0 978-1-138-05326-7.

Bludau, Kaitlyn R. (2021). "Company Activism & Black Lives Matter: A Content Analysis of BLM Instagram Posts". M.S. United States – Florida: The Florida State University. 66 pp. ISBN: 9798516064036. URL: https://www.proquest.com/docview/2547058155/abstract/185C493055C24DB4PQ/1 (visited on 10/13/2021).

Bojanowski, Piotr et al. (2016). "Enriching Word Vectors with Subword Information". Version 2. In: DOI: 10.48550/ARXIV.1607.04606. URL: https://arxiv.org/abs/1607.04606 (visited on 04/05/2023).

Bond, Robert M. et al. (Sept. 2012). "A 61-Million-Person Experiment in Social Influence and Political Mobilization". In: *Nature* 489.7415 (7415), pp. 295–298. ISSN: 1476-4687. DOI: 10.1038/nature11421. URL: https://www.nature.com/articles/nature11421 (visited on 10/12/2021).

Bourdieu, Pierre (Spr. 1989). "Social Space and Symbolic Power". In: *Sociological Theory* 7.1, p. 14. ISSN: 07352751. DOI: 10.2307/202060. JSTOR: 202060. URL: https://www.jstor.org/stable/202060?origin=crossref (visited on 07/18/2023).

— (1990). *The Logic of Practice*. Stanford, Calif: Stanford University Press. 333 pp. ISBN: 978-0-8047-1727-4.

Bradshaw, Samantha and Philip Howard (2018). "The Global Organization of Social Media Disinformation Campaigns". In: *Journal of International Affairs* 71.1.5, pp. 23–32. JSTOR: 26508115. URL: https://www.jstor.org/stable/26508115.

Branch, Jordan (Jan. 2011). "Mapping the Sovereign State: Technology, Authority, and Systemic Change". In: *International Organization* 65.1, pp. 1–36. ISSN: 0020-8183, 1531-5088. DOI: 10.1017/S0020818310000299. URL: https://www.cambridge.org/core/product/identifier/S0020818310000299/type/journal_article (visited on 08/05/2021).

Brandeis, Lewis (Mar. 21, 1932). *New State Ice Co. v. Liebmann*. URL: https://supreme.justia.com/cases/federal/us/285/262/ (visited on 03/16/2023).

Brewster, Thomas (Jan. 19, 2021). "Facebook Confirms It's Preserving Account Data Related To Capitol Hill Riot Investigations". In: *Forbes*. URL: https://www.forbes.com/sites/thomasbrewster/2021/01/19/facebook-confirms-its-preserving-account-data-related-to-capitol-hill-riot-investigations/?sh=dba32f77f660 (visited on 03/16/2023).

Bromley-Trujillo, Rebecca and John Poe (June 2020). "The Importance of Salience: Public Opinion and State Policy Action on Climate Change". In: *Journal of Public Policy* 40.2, pp. 280–304. ISSN: 0143-814X, 1469-7815. DOI: 10.1017/S0143814X18000375. URL: https://www.cambridge.org/core/product/identifier/S0143814X18000375/type/journal_article (visited on 07/28/2023).

Brown, Lester Russell (1972). *The Interdependence of Nations*. Foreign Policy Association.

Brown, Seyom (1974). *New Forces in World Politics*. Washington, DC: The Brookings Inst. 224 pp. ISBN: 978-0-8157-1117-9 978-0-8157-1118-6.

Bryson, Bethany P. (Feb. 20, 2020). "Polarizing the Middle: Internet Exposure and Public Opinion". In: *International Journal of Sociology and Social Policy* 40.1/2, pp. 99–113. ISSN: 0144-333X. DOI: 10.1108/IJSSP-09-2019-0181. URL: https://www.emerald.com/insight/content/doi/10.1108/IJSSP-09-2019-0181/full/html (visited on 08/03/2023).

Burt, Tom (May 6, 2019). *Protecting Democratic Elections through Secure, Verifiable Voting*. Microsoft On the Issues. URL: https://blogs.microsoft.com/on-the-issues/2019/05/06/protecting-democratic-elections-through-secure-verifiable-voting/ (visited on 10/12/2021).

Bush, George (July 17, 2002). *Remarks by President Bush and President Kwasniewski of Poland in An Exchange of Toasts*. Office of the Press Secretary. URL: https://georgewbush-whitehouse.archives.gov/news/releases/2002/07/text/20020717-13.html.

Byun, Joshua and Austin Carson (Dec. 20, 2022). "More than a Number: Aging Leaders in International Politics". In: *International Studies Quarterly* 67.1, sqad008. ISSN: 0020-

8833, 1468-2478. DOI: 10.1093/isq/sqad008. URL: https://academic.oup.com/isq/article/doi/10.1093/isq/sqad008/7032858 (visited on 07/17/2023).

Campante, Filipe, Ruben Durante, and Francesco Sobbrio (Aug. 1, 2018). "Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation". In: *Journal of the European Economic Association* 16.4, pp. 1094–1136. ISSN: 1542-4766. DOI: 10.1093/jeea/jvx044. URL: https://doi.org/10.1093/jeea/jvx044 (visited on 10/12/2021).

Camut, Nicolas and Ellen Boonen (Februaruy 23, 2023). "'We Are All Ukrainian.' How the Yellow-and-Blue Flag Won over Europe". In: *Politico*. URL: https://www.politico.eu/article/ukraine-russia-war-vladimyr-putin-volodymyr-zelenskyy-emmanuel-macron-yellow-and-blue-flag-won-over-europe/.

Carrière-Swallow, Yan and Vikram Haksar (2019). *The Economics and Implications of Data An Integrated Perspective*. INTERNATIONAL MONETARY FUND. ISBN: 978-1-5135-1143-6.

Carter, David B. and Randall W. Stone (2015). "Democracy and Multilateralism: The Case of Vote Buying in the UN General Assembly". In: *International Organization* 69.1, pp. 1–33. ISSN: 0020-8183, 1531-5088. DOI: 10.1017/S0020818314000186. URL: https://www.cambridge.org/core/product/identifier/S0020818314000186/type/journal_article (visited on 03/29/2023).

Carter, Joseph, Shiraz Maher, and Peter Neumann (2014). "#Greenbirds: Measuring Importance and Influence in Syrian Foreign Fighter Networks". In: *The International Centre For the Study of Radicalisation and Political Violence*.

Cave, Damien (Jan. 22, 2021). "An Australia With No Google? The Bitter Fight Behind a Drastic Threat". In: *The New York Times. Business*. ISSN: 0362-4331. URL: https://www.nytimes.com/2021/01/22/business/australia-google-facebook-news-media.html (visited on 10/12/2021).

Checkel, Jeffrey T. (Oct. 2005). "International Institutions and Socialization in Europe: Introduction and Framework". In: *International Organization* 59.04. ISSN: 0020-8183, 1531-5088. DOI: 10.1017/S0020818305050289. URL: http://www.journals.cambridge.org/abstract_S0020818305050289 (visited on 03/29/2023).

Chen, Dawn, Joshua C. Peterson, and Thomas L. Griffiths (2017). "Evaluating Vector-Space Models of Analogy". Version 2. In: DOI: 10.48550/ARXIV.1705.04416. URL: https://arxiv.org/abs/1705.04416 (visited on 07/12/2023).

Chen, Ian Tsung-yen (Apr. 27, 2022). "The Crisis of COVID-19 and the Political Economy of China's Vaccine Diplomacy". In: *Foreign Policy Analysis* 18.3, orac014. ISSN: 1743-8586, 1743-8594. DOI: 10.1093/fpa/orac014. URL: https://academic.oup.com/fpa/article/doi/10.1093/fpa/orac014/6574846 (visited on 03/29/2023).

Chen, Zenan and Xiaoge Xu (June 1, 2021). "COVID-19 News Reporting and Engaging in the Age of Social Media: Comparing Xinhua News Agency and The Paper". In: *Global Media and China* 6.2, pp. 152–170. ISSN: 2059-4364. DOI: 10.1177/20594364211017364. URL: https://doi.org/10.1177/20594364211017364 (visited on 04/17/2022).

Cloud Act Resources (Apr. 9, 2019). URL: https://www.justice.gov/dag/cloudact (visited on 10/13/2021).

Copelovitch, Mark and Stephanie Rickard (Sept. 24, 2021). "Partisan Technocrats: How Leaders Matter in International Organizations". In: *Global Studies Quarterly* 1.3, ksab021. ISSN: 2634-3797. DOI: 10.1093/isagsq/ksab021. URL: https://academic.oup.com/isagsq/article/doi/10.1093/isagsq/ksab021/6375182 (visited on 03/29/2023).

COVID-19 Community Mobility Report (2021). COVID-19 Community Mobility Report. URL: https://www.google.com/covid19/mobility?hl=en (visited on 10/12/2021).

COVID-19 Response | Google.Org (2021). URL: https://google.org/intl/en_us/covid-19/ (visited on 10/12/2021).

Criddle, Cristina (May 26, 2021). "Instagram Lets Users Hide Likes to Reduce Social Media Pressure". In: *BBC News. Technology.* URL: https://www.bbc.com/news/technology-57254488 (visited on 10/08/2021).

Dailey, Doug (June 18, 2020). "Netezza and IBM Cloud Pak for Data: A Knockout Combo for Tough Data". In: *Journey to AI Blog.* URL: https://www.ibm.com/blogs/journey-to-ai/2020/06/netezza-and-ibm-cloud-pak-a-knockout-combo-for-tough-data/.

Data Policy (Sept. 29, 2016). Meta, Inc. URL: https://www.facebook.com/about/privacy/previous (visited on 03/16/2023).

De Benedictis-Kessner, Justin et al. (Nov. 2019). "Persuading the Enemy: Estimating the Persuasive Effects of Partisan Media with the Preference-Incorporating Choice and Assignment Design". In: *American Political Science Review* 113.4, pp. 902–916. ISSN: 0003-0554, 1537-5943. DOI: 10.1017/S0003055419000418. URL: https://www.cambridge.org/core/product/identifier/S0003055419000418/type/journal_article (visited on 08/09/2023).

Deschamps, Tara (May 25, 2021). "Facebook to Pay 14 Canadian Publishers for Some News Content Posted to Its Platform". In: *Toronto Star.* URL: https://www.thestar.com/business/2021/05/25/facebook-to-pay-14-cdn-publishers-for-some-news-content-posted-to-its-platform.html (visited on 03/16/2023).

Deutsch, Karl W. (1957). *Political Community and the North American Area: International Organization in the Light of Historical Experience.* Princeton Legacy Library. Princeton, New Jersey: Princeton University Press. 227 pp. ISBN: 978-1-4008-7851-2.

Diakopoulos, Nicholas et al. (2018). "I Vote for—How Search Informs Our Choice of Candidate". In: *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple, M. Moore and D. Tambini (Eds.)* 22.

Dimitrova, Daniela V. et al. (Feb. 1, 2014). "The Effects of Digital Media on Political Knowledge and Participation in Election Campaigns: Evidence From Panel Data". In: *Communication Research* 41.1, pp. 95–118. ISSN: 0093-6502. DOI: 10.1177/0093650211426004. URL: https://doi.org/10.1177/0093650211426004 (visited on 10/12/2021).

Dosovitskiy, Alexey et al. (June 3, 2021). *An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale*. arXiv: 2010.11929 [cs]. URL: http://arxiv.org/abs/2010.11929 (visited on 06/20/2023). preprint.

Duque, Marina G (Sept. 1, 2018). "Recognizing International Status: A Relational Approach". In: *International Studies Quarterly* 62.3, pp. 577–592. ISSN: 0020-8833, 1468-2478. DOI: 10.1093/isq/sqy001. URL: https://academic.oup.com/isq/article/62/3/577/4962448 (visited on 03/29/2023).

Durkin, Kevin and Gina Conti-Ramsden (Sept. 2007). "Language, Social Behavior, and the Quality of Friendships in Adolescents With and Without a History of Specific Language Impairment". In: *Child Development* 78.5, pp. 1441–1457. ISSN: 0009-3920, 1467-8624. DOI: 10.1111/j.1467-8624.2007.01076.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1467-8624.2007.01076.x (visited on 08/16/2023).

Edwards, John (2009). *Language and Identity: An Introduction*. Key Topics in Sociolinguistics. Cambridge, UK ; New York: Cambridge University Press. 314 pp. ISBN: 978-0-521-87381-9 978-0-521-69602-9.

Efron, Bradley (May 1, 2003). "Second Thoughts on the Bootstrap". In: *Statistical Science* 18.2. ISSN: 0883-4237. DOI: 10.1214/ss/1063994968. URL: https://projecteuclid.org/journals/statistical-science/volume-18/issue-2/Second-Thoughts-on-the-Bootstrap/10.1214/ss/1063994968.full (visited on 03/29/2023).

Eisenman, Joshua (Mar. 16, 2023). *China's Media Propaganda in Africa: A Strategic Assessment*. Unitted States Institute of Peace. URL: https://www.usip.org/publications/2023/03/chinas-media-propaganda-africa-strategic-assessment.

El Damanhoury, Kareem and Nisha Garud-Paktar (Aug. 31, 2021). "Soft Power Journalism: A Visual Framing Analysis of COVID-19 on Xinhua and VOA's Instagram Pages". In: *Digital Journalism* 0.0, pp. 1–23. ISSN: 2167-0811. DOI: 10.1080/21670811.2021.1957969. URL: https://doi.org/10.1080/21670811.2021.1957969 (visited on 04/17/2022).

Epstein, Robert and Ronald E. Robertson (Aug. 18, 2015). "The Search Engine Manipulation Effect (SEME) and Its Possible Impact on the Outcomes of Elections". In: *Proceedings*

of the National Academy of Sciences 112.33, E4512–E4521. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1419828112. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1419828112 (visited on 10/22/2021).

Epstein, Robert, Ronald E. Robertson, et al. (2017). "Suppressing the Search Engine Manipulation Effect (SEME)". In: *Proceedings of the ACM on Human-Computer Interaction* 1 (CSCW), pp. 1–22.

Facebook (2021). *How Do I Report a Child under the Age of 13 on Facebook?* Facebook Help Center. URL: https://www.facebook.com/help/157793540954833 (visited on 10/03/2021).

Facebook Data For Good Social Connectedness Index (2021). URL: https://dataforgood.facebook.com/dfg/tools/social-connectedness-index (visited on 10/08/2021).

Farrell, Henry and Abraham L. Newman (July 2019). "Weaponized Interdependence: How Global Economic Networks Shape State Coercion". In: *International Security* 44.1, pp. 42–79. ISSN: 0162-2889, 1531-4804. DOI: 10.1162/isec_a_00351. URL: https://www.mitpressjournals.org/doi/abs/10.1162/isec_a_00351 (visited on 12/05/2020).

Fearon, James D. (1995). "Rationalist Explanations for War". In: *International Organization* 49.3, pp. 379–414. ISSN: 0020-8183, 1531-5088. DOI: 10.1017/S0020818300033324. URL: https://www.cambridge.org/core/product/identifier/S0020818300033324/type/journal_article (visited on 08/09/2023).

Fischer, Sean, Kokil Jaidka, and Yphtach Lelkes (2020). "Auditing Local News Presence on Google News". In: *Nature Human Behaviour* 4.12, pp. 1236–1244.

Flyverbom, Mikkel and John Murray (July 2018). "Datastructuring—Organizing and Curating Digital Traces into Action". In: *Big Data & Society* 5.2, p. 205395171879911. ISSN: 2053-9517, 2053-9517. DOI: 10.1177/2053951718799114. URL: http://journals.sagepub.com/doi/10.1177/2053951718799114 (visited on 10/13/2021).

Foer, Franklin (2017). *World without Mind: The Existential Threat of Big Tech*. New York: Penguin Press. 257 pp. ISBN: 978-1-101-98111-5.

Foreign Relations, Chicago Council on (2021). *American Public Opinion and United States Foreign Policy Series*. Inter-university Consortium for Political and Social Research [distributor]. URL: https://www.icpsr.umich.edu/web/ICPSR/series/4.

Form 10-K (Dec. 31, 2022). Meta Platforms, Inc. URL: https://d18rn0p25nwr6d.cloudfront.net/CIK-0001326801/e574646c-c642-42d9-9229-3892b13aabfb.pdf.

Franceschi-Bicchierai, Lorenzo (June 10, 2020). "Facebook Helped the FBI Hack a Child Predator". In: *Vice*. URL: https://www.vice.com/en/article/v7gd9b/facebook-helped-fbi-hack-child-predator-buster-hernandez (visited on 03/16/2023).

Frank, John (May 3, 2019). *Taking Further Steps to Support Electoral Integrity in Europe*. EU Policy Blog. URL: https://blogs.microsoft.com/eupolicy/2019/05/03/supporting-european-electoral-integrity/ (visited on 10/12/2021).

Frankovic, Kathy and Taylor Orth (Feb. 16, 2023). "A Growing Share of Americans View China as an Enemy of the United States". In: *YouGovAmerica*. URL: https://today.yougov.com/topics/international/articles-reports/2023/02/16/growing-share-americans-view-china-enemy-us.

Freund, Caroline L and Diana Weinhold (Jan. 1, 2004). "The Effect of the Internet on International Trade". In: *Journal of International Economics* 62.1, pp. 171–189. ISSN: 0022-1996. DOI: 10.1016/S0022-1996(03)00059-X. URL: https://www.sciencedirect.com/science/article/pii/S002219960300059X (visited on 10/12/2021).

Gartzke, Erik (June 2000). "Preferences and the Democratic Peace". In: *International Studies Quarterly* 44.2, pp. 191–212. ISSN: 0020-8833, 1468-2478. DOI: 10.1111/0020-8833.00155. URL: https://academic.oup.com/isq/article-lookup/doi/10.1111/0020-8833.00155 (visited on 08/16/2023).

Geist, Edward and Andrew Lohn (2018). *How Might Artificial Intelligence Affect the Risk of Nuclear War?* RAND Corporation. DOI: 10.7249/PE296. URL: https://www.rand.org/pubs/perspectives/PE296.html (visited on 02/16/2021).

Gentzkow, Matthew, Jesse M. Shapiro, and Matt Taddy (2019). "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech". In: *Econometrica* 87.4, pp. 1307–1340. ISSN: 0012-9682. DOI: 10.3982/ECTA16566. URL: https://www.econometricsociety.org/doi/10.3982/ECTA16566 (visited on 03/29/2023).

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer (Feb. 2008). "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment". In: *American Political Science Review* 102.1, pp. 33–48. ISSN: 0003-0554, 1537-5943. DOI: 10.1017/S000305540808009X. URL: https://www.cambridge.org/core/product/identifier/S000305540808009X/type/journal_article (visited on 10/12/2021).

Gerstenzang, James (Mar. 9, 1999). "Clinton Reaches Out to Ravaged Nicaragua / President Begins 4-Day Central America Tour". In: *Los Angeles Times*. URL: https://www.sfgate.com/news/article/Clinton-Reaches-Out-to-Ravaged-Nicaragua-2943137.php.

Ghose, Anindya, Avi Goldfarb, and Sang Pil Han (2013). "How Is the Mobile Internet Different? Search Costs and Local Activities". In: *Information Systems Research* 24.3, pp. 613–631.

Gibler, Douglas M. (2009). *International Military Alliances, 1648-2008*. Correlates of War Series. Washington, D.C: CQ Press. 2 pp. ISBN: 978-1-56802-824-8.

Gil de Zúñiga, Homero, Logan Molyneux, and Pei Zheng (Aug. 1, 2014). "Social Media, Political Expression, and Political Participation: Panel Analysis of Lagged and Concurrent Relationships". In: *Journal of Communication* 64.4, pp. 612–634. ISSN: 0021-9916. DOI: 10.1111/jcom.12103. URL: https://doi.org/10.1111/jcom.12103 (visited on 10/12/2021).

Gillespie, Tarleton (2017). "Algorithmically Recognizable: Santorum's Google Problem, and Google's Santorum Problem". In: *Information, communication & society* 20.1, pp. 63–80.

Glaser, Charles L. (Mar. 1, 1996). "Realists as Optimists: Cooperation as Self-help". In: *Security Studies* 5.3, pp. 122–163. ISSN: 0963-6412. DOI: 10.1080/09636419608429278. URL: https://doi.org/10.1080/09636419608429278 (visited on 10/07/2021).

— (2010). *Rational Theory of International Politics: The Logic of Competition and Cooperation*. Princeton, N.J: Princeton University Press. 314 pp. ISBN: 978-0-691-14371-2 978-0-691-14372-9.

Gleicher, Nathaniel (Nov. 13, 2018). *How Do We Work With Our Partners to Combat Information Operations?* About Facebook. URL: https://about.fb.com/news/2018/11/investigating-threats/ (visited on 10/12/2021).

Golebiewski, Michael and Danah Boyd (2019). *Data Voids: Where Missing Data Can Easily Be Exploited*. Data & Society Research Institute.

Google | Protect Your Election | Protection from Digital Attacks (2021). Google | Protect Your Election | Protection from digital attacks. URL: https://protectyourelection.withgoogle.com/intl/en (visited on 10/12/2021).

Google Organic CTR History (Mar. 2022). Advanced Web Ranking. URL: https://www.advancedwebranking.com/ctrstudy/ (visited on 04/07/2022).

Google Search API (2022). SerpAPI. URL: https://serpapi.com/ (visited on 10/01/2022).

Google's Revolving Door (US) (Apr. 26, 2016). Tech Transparency Project. URL: https://www.techtransparencyproject.org/articles/googles-revolving-door-us (visited on 10/12/2021).

Gouvard, Paul, Amir Goldberg, and Sameer B. Srivastava (June 15, 2023). "Doing Organizational Identity: Earnings Surprises and the Performative Atypicality Premium". In: *Administrative Science Quarterly*, p. 00018392231180872. ISSN: 0001-8392, 1930-3815. DOI: 10.1177/00018392231180872. URL: http://journals.sagepub.com/doi/10.1177/00018392231180872 (visited on 07/17/2023).

Granka, Laura A. (2010). "The Politics of Search: A Decade Retrospective". In: *The Information Society* 26.5, pp. 364–374.

Green, Joshua (May 31, 2013). *Google's Eric Schmidt Invests in Obama's Big Data Brains*. Bloomberg. URL: https://www.bloomberg.com/news/articles/2013-05-30/googles-eric-schmidt-invests-in-obamas-big-data-brains (visited on 10/12/2021).

Greenwald, Glenn (May 13, 2014). *No Place to Hide: Edward Snowden, the NSA, and the U.S. Surveillance State*. Macmillan. 273 pp. ISBN: 978-1-62779-073-4.

Grossman, Gene and Elhanan Helpman (Aug. 1992). *Protection For Sale*. w4149. Cambridge, MA: National Bureau of Economic Research, w4149. DOI: 10.3386/w4149. URL: http://www.nber.org/papers/w4149.pdf (visited on 08/10/2023).

Grossman, Gene and Esteban Rossi-Hansberg (2006). "The Rise of Offshoring: It's Not Wine for Cloth Anymore". In: *Jackson Hole Economic Symposium*, pp. 59–102.

Haag, Matthew (Feb. 4, 2019). "FamilyTreeDNA Admits to Sharing Genetic Data With F.B.I." In: *The New York Times*. URL: https://www.nytimes.com/2019/02/04/business/family-tree-dna-fbi.html?login=smartlock&auth=login-smartlock (visited on 03/16/2023).

Hall, Madison (Jan. 14, 2021). "Parler Turned over Identifying Information to the FBI after a User Threatened to Kill Trump and Supreme Court Justice Amy Coney Barrett". In: *Business Insider*. URL: https://www.businessinsider.com/parler-sent-fbi-user-info-after-threat-to-kill-trump-2021-1 (visited on 03/16/2023).

Hamilton, Antonia F. de C. and Frida Lind (2016). "Audience Effects: What Can They Tell Us about Social Neuroscience, Theory of Mind and Autism?" In: *Culture and Brain* 4.2, pp. 159–177. ISSN: 2193-8652. DOI: 10.1007/s40167-016-0044-5. pmid: 27867833. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5095155/ (visited on 10/13/2021).

Hannák, Anikó et al. (June 15, 2017). "Measuring Personalization of Web Search". arXiv: 1706.05011 [cs]. URL: http://arxiv.org/abs/1706.05011 (visited on 11/28/2021).

Harris, Mark (Sept. 30, 2021). "How a Secret Google Geofence Warrant Helped Catch the Capitol Riot Mob". In: *Wired*. URL: https://www.wired.com/story/capitol-riot-google-geofence-warrant/ (visited on 03/16/2023).

Harris, Zellig S. (Aug. 1954). "Distributional Structure". In: *WORD* 10.2-3, pp. 146–162. ISSN: 0043-7956, 2373-5112. DOI: 10.1080/00437956.1954.11659520. URL: http://www.tandfonline.com/doi/full/10.1080/00437956.1954.11659520 (visited on 04/05/2023).

Heller, Monica (Nov. 2003). "Globalization, the New Economy, and the Commodification of Language and Identity". In: *Journal of Sociolinguistics* 7.4, pp. 473–492. ISSN: 13606441. DOI: 10.1111/j.1467-9841.2003.00238.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9841.2003.00238.x (visited on 08/16/2023).

Hemmer, Christopher and Peter J. Katzenstein (2002). "Why Is There No NATO in Asia? Collective Identity, Regionalism, and the Origins of Multilateralism". In: *International Organization* 56.3, pp. 575–607. ISSN: 0020-8183, 1531-5088. DOI: 10.1162/002081802760199890. URL: https://www.cambridge.org/core/product/identifier/S002081830244180X/type/journal_article (visited on 03/29/2023).

Heymans, Maureen and Murali Viswanathan (n.d.). *Introducing Google Social Search: I Finally Found My Friend's New York Blog!* URL: https://googleblog.blogspot.com/2009/10/introducing-google-social-search-i.html.

Horling, Bryan and Matthew Kulick (Dec. 4, 2009). *Personalized Search for Everyone*. Google Official Blog. URL: https://googleblog.blogspot.com/2009/12/personalized-search-for-everyone.html (visited on 10/01/2021).

How Google Cloud Is Helping during Coronavirus (2021). Google Cloud Blog. URL: https://cloud.google.com/blog/topics/inside-google-cloud/how-google-cloud-is-helping-during-covid-19/ (visited on 10/12/2021).

How Google Determines the Language of Search Results (n.d.). Google. URL: https://support.google.com/websearch/answer/13511324?hl=en#zippy=.

How Google Protects Your Privacy & Keeps You in Control (2022). Google Chat Help. URL: https://support.google.com/chat/answer/10400210 (visited on 04/07/2022).

Howison, James et al. (Dec. 2011). "Validity Issues in the Use of Social Network Analysis with Digital Trace Data". In: *Journal of the Association for Information Systems* 12.12, pp. 767–797. ISSN: 15369323. DOI: 10.17705/1jais.00282. URL: http://aisel.aisnet.org/jais/vol12/iss12/2/ (visited on 10/03/2021).

Hu, Desheng et al. (2019). "Auditing the Partisanship of Google Search Snippets". In: *The World Wide Web Conference on - WWW '19*. The World Wide Web Conference. San Francisco, CA, USA: ACM Press, pp. 693–704. ISBN: 978-1-4503-6674-8. DOI: 10.1145/3308558.3313654. URL: http://dl.acm.org/citation.cfm?doid=3308558.3313654 (visited on 04/09/2022).

Human Rights Impact Assessment (Oct. 2018). Business for Social Responsibility. URL: https://about.fb.com/wp-content/uploads/2018/11/bsr-facebook-myanmar-hria_final.pdf.

Hyde, Susan D. and Elizabeth N. Saunders (2020). "Recapturing Regime Type in International Relations: Leaders, Institutions, and Agency Space". In: *International Organization* 74.2, pp. 363–395. ISSN: 0020-8183, 1531-5088. DOI: 10.1017/S0020818319000365. URL: https://www.cambridge.org/core/product/identifier/S0020818319000365/type/journal_article (visited on 08/09/2023).

Introna, Lucas and Helen Nissenbaum (Oct. 28, 2000). "Shaping The Web: Why The Politics Of Search Engines Matters". In: *The Information Society* 16, pp. 169–185. DOI: 10.1080/01972240050133634.

Investigation, FBI – Federal Bureau of (Feb. 23, 2021). URL: https://www.facebook.com/FBI/photos/the-fbi-is-still-seeking-information-on-people-who-took-part-in-the-violence-at-/10158937641066212/.

Isaac, Mike et al. (Feb. 17, 2021). "Facebook Blocks News in Australia, Diverging With Google on Proposed Law". In: *The New York Times. Technology*. ISSN: 0362-4331. URL: https://www.nytimes.com/2021/02/17/technology/facebook-google-australia-news.html (visited on 10/12/2021).

Isin, Engin and Evelyn Sharon Ruppert (2019). "Data's Empire: Postcolonial Data Politics". In: *Data Politics: Worlds, Subjects, Rights*. Routledge Studies in International Political Sociology. London ; New York: Routledge, Taylor & Francis Group.

Jennings, Rebecca (June 3, 2020). *Who Are the Black Squares and Cutesy Illustrations Really For?* Vox. URL: https://www.vox.com/the-goods/2020/6/3/21279336/blackout-tuesday-black-lives-matter-instagram-performative-allyship (visited on 10/13/2021).

Jervis, Robert (1976). *Perception and Misperception in International Politics: New Edition*. REV - Revised. Princeton University Press. ISBN: 978-0-691-17743-4. DOI: 10.2307/j.ctvc77bx3. JSTOR: j.ctvc77bx3. URL: https://www.jstor.org/stable/j.ctvc77bx3 (visited on 10/08/2021).

— (Oct. 2, 1990). *The Meaning of the Nuclear Revolution: Statecraft and the Prospect of Armageddon*. Reprint edition. Ithaca: Cornell University Press. 272 pp. ISBN: 978-0-8014-9565-6.

— (Dec. 1, 1997). "Complexity and the Analysis of Political and Social Life". In: *Political Science Quarterly* 112.4, pp. 569–593. ISSN: 0032-3195, 1538-165X. DOI: 10.2307/2657692. URL: https://academic.oup.com/psq/article/112/4/569/7043671 (visited on 04/04/2023).

Jiang, Min (2014). "The Business and Politics of Search Engines: A Comparative Study of Baidu and Google's Search Results of Internet Events in China". In: *New media & society* 16.2, pp. 212–233.

Johnson, Kevin (2021). *NSA Taps Data from 9 Major Net Firms*. USATODAY. URL: https://www.usatoday.com/story/news/2013/06/06/nsa-surveillance-internet-companies/2398345/ (visited on 10/13/2021).

Joint Press Conference With President Barack Obama And President Felipe Calderon Of Mexico (Apr. 16, 2009). The White House Office of the Press Secretary. URL: https:

//obamawhitehouse.archives.gov/the-press-office/joint-press-conference-with-president-barack-obama-and-president-felipe-calderon-me (visited on 03/28/2023).

Just, Natascha and Michael Latzer (Mar. 2017). "Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet". In: *Media, Culture & Society* 39.2, pp. 238–258. ISSN: 0163-4437, 1460-3675. DOI: 10.1177/0163443716643157. URL: http://journals.sagepub.com/doi/10.1177/0163443716643157 (visited on 07/03/2023).

Kalinowski, Alexander and Yuan An (2020). "A Survey of Embedding Space Alignment Methods for Language and Knowledge Graphs". Version 1. In: DOI: 10.48550/ARXIV.2010.13688. URL: https://arxiv.org/abs/2010.13688 (visited on 04/03/2023).

Kamvar, Sp (June 28, 2005). *Search Gets Personal*. Google Official Blog. URL: https://googleblog.blogspot.com/2005/06/search-gets-personal.html (visited on 10/01/2021).

Kay, Matthew, Cynthia Matuszek, and Sean A. Munson (2015). "Unequal Representation and Gender Stereotypes in Image Search Results for Occupations". In: *Proceedings of the 33rd Annual Acm Conference on Human Factors in Computing Systems*, pp. 3819–3828.

Kaye, Byron (n.d.). "Facebook 'unfriends' Australia: Uproar as News Pages Go Dark". In: *Reuters* (). URL: https://www.reuters.com/article/us-australia-media-facebook-idUSKBN2AI02A.

Keck, Margaret E. and Kathryn Sikkink (1998). *Activists beyond Borders: Advocacy Networks in International Politics*. Ithaca, N.Y: Cornell University Press. 228 pp. ISBN: 978-0-8014-3444-0 978-0-8014-8456-8.

Keohane, Robert O. (1984). *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton University Press. 307 pp. ISBN: 978-1-4008-2026-9. Google Books: HnvpdocqT9EC.

Keohane, Robert O. and Joseph S. Nye (2012). *Power and Interdependence*. 4th ed. Boston: Longman. 330 pp. ISBN: 978-0-205-08291-9.

Kim, Sei-Hill (July 1, 2008). "Testing the Knowledge Gap Hypothesis in South Korea: Traditional News Media, the Internet, and Political Learning". In: *International Journal of Public Opinion Research* 20.2, pp. 193–210. ISSN: 0954-2892. DOI: 10.1093/ijpor/edn019. URL: https://doi.org/10.1093/ijpor/edn019 (visited on 10/12/2021).

Kim, Yonghwan, Hsuan-Ting Chen, and Homero Gil De Zúñiga (Nov. 2013). "Stumbling upon News on the Internet: Effects of Incidental News Exposure and Relative Entertainment Use on Political Engagement". In: *Computers in Human Behavior* 29.6, pp. 2607–2614. ISSN: 07475632. DOI: 10.1016/j.chb.2013.06.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S0747563213001970 (visited on 08/03/2023).

Kindleberger, Charles P. (Spr. 1969). "American Business Abroad". In: *The International Executive* 11.2, pp. 11–12. ISSN: 00206652, 1522709X. DOI: 10.1002/tie.5060110207. URL: https://onlinelibrary.wiley.com/doi/10.1002/tie.5060110207 (visited on 03/16/2023).

King, Gary, Jennifer Pan, and Margaret E. Roberts (2013). "How Censorship in China Allows Government Criticism but Silences Collective Expression". In: *American Political Science Review* 107.02, pp. 326–343. URL: http://journals.cambridge.org/abstract_S0003055413000014 (visited on 03/16/2015).

Kliman-Silver, Chloe et al. (2015). "Location, Location, Location: The Impact of Geolocation on Web Search Personalization". In: *Proceedings of the 2015 Internet Measurement Conference*, pp. 121–127.

Knobloch-Westerwick, Silvia et al. (2015). "Science Exemplars in the Eye of the Beholder: How Exposure to Online Science Information Affects Attitudes". In: *Science Communication* 37.5, pp. 575–601.

Knorre, Aleksei, Ruslan Kuchakov, and Dmitriy Skougarevskiy (2023). "Stakeholder Activism and Foreign Firm Exit From Russia in 2022". In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.4471582. URL: https://www.ssrn.com/abstract=4471582 (visited on 08/10/2023).

Kono, Daniel Y. (Aug. 2008). "Does Public Opinion Affect Trade Policy?" In: *Business and Politics* 10.2, pp. 1–19. ISSN: 1369-5258, 1469-3569. DOI: 10.2202/1469-3569.1224. URL: https://www.cambridge.org/core/product/identifier/S1369525800003557/type/journal_article (visited on 07/28/2023).

Kozlowski, Austin C., Matt Taddy, and James A. Evans (Oct. 2019). "The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings". In: *American Sociological Review* 84.5, pp. 905–949. ISSN: 0003-1224, 1939-8271. DOI: 10.1177/0003122419877135. URL: http://journals.sagepub.com/doi/10.1177/0003122419877135 (visited on 09/07/2021).

Krasner, Stephen D. (Apr. 1976). "State Power and the Structure of International Trade". In: *World Politics* 28.3, pp. 317–347. ISSN: 0043-8871, 1086-3338. DOI: 10.2307/2009974. URL: https://www.cambridge.org/core/product/identifier/S0043887100021055/type/journal_article (visited on 03/16/2023).

— (1978). *Defending the National Interest: Raw Materials Investments and U.S. Foreign Policy*. In collab. with Harvard University. Princeton Paperbacks. Princeton, N.J: Princeton Univ. Press. 404 pp. ISBN: 978-0-691-02182-9 978-0-691-07600-3.

Kreps, Sarah E (2020). *Social Media and International Relations*. ISBN: 978-1-108-92037-7. URL: https://doi.org/10.1017/9781108920377 (visited on 02/16/2021).

Kulshrestha, Juhi et al. (2017). "Quantifying Search Bias: Investigating Sources of Bias for Political Searches in Social Media". In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 417–432.

Laidlaw, Emily B. (2009). "Private Power, Public Interest: An Examination of Search Engine Accountability". In: *International Journal of Law and Information Technology* 17.1, pp. 113–145.

Lawrence, Steve and C Lee Giles (1999). "Accessibility of Information on the Web". In: 400, p. 3.

Lee, Melissa M. and Nan Zhang (Jan. 1, 2017). "Legibility and the Informational Foundations of State Capacity". In: *The Journal of Politics* 79.1, pp. 118–132. ISSN: 0022-3816. DOI: 10.1086/688053. URL: https://www.journals.uchicago.edu/doi/full/10.1086/688053 (visited on 10/12/2021).

Levy, Frank S. (2017). "Computers and Populism". In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.3091867. URL: https://www.ssrn.com/abstract=3091867 (visited on 08/09/2023).

Lieber, Keir A. and Daryl G. Press (Apr. 2017). "The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence". In: *International Security* 41.4, pp. 9–49. ISSN: 0162-2889, 1531-4804. DOI: 10.1162/ISEC_a_00273. URL: https://direct.mit.edu/isec/article/41/4/9-49/12158 (visited on 10/07/2021).

Liedke, Jacob and Katerina Eva Matsa (Sept. 20, 2022). *Social Media and News Fact Sheet*. Pew Research Center. URL: https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/ (visited on 03/16/2023).

Luo, Queenie, Michael J. Puett, and Michael D. Smith (2023). "A Perspectival Mirror of the Elephant: Investigating Language Bias on Google, ChatGPT, Wikipedia, and YouTube". Version 2. In: DOI: 10.48550/ARXIV.2303.16281. URL: https://arxiv.org/abs/2303.16281 (visited on 06/29/2023).

Lyon, David (Oct. 19, 2015). *Surveillance After Snowden*. John Wiley & Sons. 111 pp. ISBN: 978-0-7456-9088-9. Google Books: gBbICgAAQBAJ.

Maas, Paige (July 25, 2019). "Facebook Disaster Maps: Aggregate Insights for Crisis Response & Recovery". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. KDD '19: The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Anchorage AK USA: ACM, pp. 3173–3173. ISBN: 978-1-4503-6201-6. DOI: 10.1145/3292500.3340412. URL: https://dl.acm.org/doi/10.1145/3292500.3340412 (visited on 10/12/2021).

Maddalena Conte, Pierre Cotterlaz, and Thierry Mayer (2022). "The CEPII Gravity Database". In: *CEPII Research Center* (Working Papers 2022-05).

Marshall, Monty (Apr. 23, 2020). "Polity5: Political Regime Characteristics and Transitions, 1800-2018". In: *Center for Systemic Peace* 5.

McInnes, Leland, John Healy, and James Melville (Sept. 17, 2020). "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction". arXiv: 1802.03426 [cs, stat]. URL: http://arxiv.org/abs/1802.03426 (visited on 04/14/2022).

McPherson, Miller, Lynn Smith-Lovin, and James M Cook (Aug. 2001). "Birds of a Feather: Homophily in Social Networks". In: *Annual Review of Sociology* 27.1, pp. 415–444. ISSN: 0360-0572, 1545-2115. DOI: 10.1146/annurev.soc.27.1.415. URL: https://www.annualreviews.org/doi/10.1146/annurev.soc.27.1.415 (visited on 06/28/2023).

Mearsheimer, John J. (2018). *The Great Delusion: Liberal Dreams and International Realities*. The Henry l. Stimson Lectures Series. New Haven, CT: Yale University Press. ISBN: 978-0-300-23419-0.

Mejova, Yelena, Tatiana Gracyk, and Ronald E. Robertson (Feb. 23, 2022). "Googling for Abortion: Search Engine Mediation of Abortion Accessibility in the United States". In: *Journal of Quantitative Description: Digital Media* 2. ISSN: 2673-8813. DOI: 10.51685/jqd.2022.007. arXiv: 2202.11760. URL: http://arxiv.org/abs/2202.11760 (visited on 03/22/2022).

Meta, Inc. (2023a). *Case Studies*. Facebook Transparency Center. URL: https://transparency.fb.com/data/government-data-requests/case-studies/ (visited on 03/16/2023).

— (2023b). *Requests by Country*. Facebook Transparency Center. URL: https://transparency.fb.com/data/government-data-requests/country/ (visited on 03/16/2023).

Meta Platforms, Inc. Form 10-Q (Mar. 31, 2023). URL: https://d18rn0p25nwr6d.cloudfront.net/CIK-0001326801/5fb5d0ea-c2c7-46f0-a26a-656df1673fac.pdf.

Metaxa, Danaë et al. (2019). "Search Media and Elections: A Longitudinal Investigation of Political Search Results". In: *Proceedings of the ACM on Human-Computer Interaction* 3 (CSCW), pp. 1–17.

Metaxas, P. Takis and Eni Mustafaraj (2009). "The Battle for the 2008 US Congressional Elections on the Web". In: *Proceedings of the 2009 WebScience: Society On-Line Conference*.

Metaxas, P. Takis and Yada Pruksachatkun (2017). "Manipulation of Search Engine Results during the 2016 US Congressional Elections". In.

"Microsoft, Amazon and Google Account for Over Half of Today's 600 Hyperscale Data Centers" (Jan. 26, 2021). In: *Synergy Research Group*. URL: https://www.srgresearch.com/articles/microsoft-amazon-and-google-account-for-over-half-of-todays-600-hyperscale-data-centers.

Mikolov, Tomas, Kai Chen, et al. (2013). "Efficient Estimation of Word Representations in Vector Space". Version 3. In: DOI: 10.48550/ARXIV.1301.3781. URL: https://arxiv.org/abs/1301.3781 (visited on 03/29/2023).

Mikolov, Tomas, Ilya Sutskever, et al. (2013). "Distributed Representations of Words and Phrases and Their Compositionality". Version 1. In: DOI: 10.48550/ARXIV.1310.4546. URL: https://arxiv.org/abs/1310.4546 (visited on 07/12/2023).

Milbauer, Jeremiah, Adarsh Mathew, and James Evans (2021). "Aligning Multidimensional Worldviews and Discovering Ideological Differences". In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, pp. 4832–4845. DOI: 10.18653/v1/2021.emnlp-main.396. URL: https://aclanthology.org/2021.emnlp-main.396 (visited on 04/06/2023).

Miller, Claire (2021). *Tech Companies Concede to Surveillance Program - The New York Times*. URL: https://www.nytimes.com/2013/06/08/technology/tech-companies-bristling-concede-to-government-surveillance-efforts.html?ref=global-home&_r=1&&pagewanted=all (visited on 10/13/2021).

Minow, Martha (2021). *Saving the News: Why the Constitution Calls for Government Action to Preserve Freedom of Speech*. Inalienable Rights Series. New York: Oxford University Press. 1 p. ISBN: 978-0-19-094844-3 978-0-19-094843-6.

Mitts, Tamar (2019). "From Isolation to Radicalization: Anti-Muslim Hostility and Support for ISIS in the West". In: *American Political Science Review* 113.1, pp. 173–94.

Mohsin, Maryam (Jan. 2, 2022). *10 Google Search Statistics You Need to Know in 2022*. Oberlo. URL: https://www.oberlo.com/blog/google-search-statistics (visited on 04/07/2022).

Mowshowitz, Abbe and Akira Kawaguchi (Sept. 1, 2005). "Measuring Search Engine Bias". In: *Information Processing & Management* 41.5, pp. 1193–1205. ISSN: 0306-4573. DOI: 10.1016/j.ipm.2004.05.005. URL: https://www.sciencedirect.com/science/article/pii/S0306457304000615 (visited on 04/09/2022).

Mozur, Paul (Aug. 20, 2021). "To Protect Users, Facebook Says It's Hiding Friends Lists on Accounts in Afghanistan." In: *The New York Times. World*. ISSN: 0362-4331. URL: https://www.nytimes.com/2021/08/20/world/asia/afghanistan-facebook.html (visited on 10/13/2021).

Nayak, Pandu (Oct. 25, 2019). *Understanding Searches Better than Ever Before*. Google. URL: https://blog.google/products/search/search-language-understanding-bert/ (visited on 10/12/2021).

Nelson, Laura K. (Oct. 2021). "Leveraging the Alignment between Machine Learning and Intersectionality: Using Word Embeddings to Measure Intersectional Experiences of the Nineteenth Century U.S. South". In: *Poetics* 88, p. 101539. ISSN: 0304422X. DOI: 10.1016/j.poetic.2021.101539. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304422X21000115 (visited on 03/29/2023).

Neutze, Jan (May 6, 2019). *Protecting Political Campaigns from Hacking*. Microsoft On the Issues. URL: https://blogs.microsoft.com/on-the-issues/2019/05/06/protecting-political-campaigns-from-hacking/ (visited on 10/12/2021).

Noble, Safiya Umoja (2013). "Google Search: Hyper-visibility as a Means of Rendering Black Women and Girls Invisible". In: *InVisible Culture* 19.

— (Feb. 20, 2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. Illustrated edition. New York: NYU Press. 256 pp. ISBN: 978-1-4798-3724-3.

Ochigame, Rodrigo and Katherine Ye (June 28, 2021). "Search Atlas: Visualizing Divergent Search Results Across Geopolitical Borders". In: *Designing Interactive Systems Conference 2021*. DIS '21. New York, NY, USA: Association for Computing Machinery, pp. 1970–1983. ISBN: 978-1-4503-8476-6. DOI: 10.1145/3461778.3462032. URL: https://doi.org/10.1145/3461778.3462032 (visited on 10/12/2021).

Ognyanova, Katherine et al. (June 2, 2020). "Misinformation in Action: Fake News Exposure Is Linked to Lower Trust in Media, Higher Trust in Government When Your Side Is in Power". In: *Harvard Kennedy School Misinformation Review*. DOI: 10.37016/mr-2020-024. URL: https://misinforeview.hks.harvard.edu/?p=1689 (visited on 02/16/2021).

Olson, Emily (Dec. 17, 2022). "A $1.6 Billion Lawsuit Alleges Facebook's Inaction Fueled Violence in Ethiopia". In: *NPR*. URL: https://www.npr.org/2022/12/17/1142873282/facebook-meta-lawsuit-ethiopia-kenya-abrham-amare.

Ørmen, Jacob (2016). "Googling the News: Opportunities and Challenges in Studying News Events through Google Search". In: *Digital Journalism* 4.1, pp. 107–124.

Our Steps to Protect State Elections in India (Mar. 31, 2021). About Facebook. URL: https://about.fb.com/news/2021/03/steps-to-protect-elections-india/ (visited on 10/12/2021).

Palmer, Glenn et al. (July 2022). "The MID5 Dataset, 2011–2014: Procedures, Coding Rules, and Description". In: *Conflict Management and Peace Science* 39.4, pp. 470–482. ISSN: 0738-8942, 1549-9219. DOI: 10.1177/0738894221995743. URL: http://journals.sagepub.com/doi/10.1177/0738894221995743 (visited on 03/30/2023).

Pan, Bing et al. (2007). "In Google We Trust: Users' Decisions on Rank, Position, and Relevance". In: *Journal of computer-mediated communication* 12.3, pp. 801–823.

Pariser, Eli (May 12, 2011). *The Filter Bubble: What The Internet Is Hiding From You*. Penguin Books Limited. 246 pp. ISBN: 978-0-14-196992-3.

Pichai, Sundar (Mar. 27, 2020). *COVID-19: $800+ Million to Support Small Businesses and Crisis Response*. Google. URL: https://blog.google/inside-google/company-announcements/commitment-support-small-businesses-and-crisis-response-covid-19/ (visited on 10/12/2021).

Placek, Matthew (Nov. 16, 2020). "Learning Democracy Digitally? The Internet and Knowledge of Democracy in Nondemocracies". In: *Democratization* 27.8, pp. 1413–1435. ISSN: 1351-0347. DOI: 10.1080/13510347.2020.1795640. URL: https://doi.org/10.1080/13510347.2020.1795640 (visited on 10/12/2021).

Pouliot, Vincent (Sept. 2007). "Pacification Without Collective Identification: Russia and the Transatlantic Security Community in the Post-Cold War Era". In: *Journal of Peace Research* 44.5, pp. 605–622. ISSN: 0022-3433, 1460-3578. DOI: 10.1177/0022343307080858. URL: http://journals.sagepub.com/doi/10.1177/0022343307080858 (visited on 03/29/2023).

Preibusch, Sören (Apr. 23, 2015). "Privacy Behaviors after Snowden". In: *Communications of the ACM* 58.5, pp. 48–55. ISSN: 0001-0782, 1557-7317. DOI: 10.1145/2663341. URL: https://dl.acm.org/doi/10.1145/2663341 (visited on 02/17/2021).

Promoting Public Safety, Privacy, and the Rule of Law Around the World: The Purpose and Impact of the CLOUD Act (Apr. 2019). U.S. Department of Justice.

Rahman, K. Sabeel (2018). "The New Utilities: Private Power, Social Infrastructure, and the Revival of the Public Utility Concept". In: *Cardozo Law Review* 39.5, pp. 1621–1692.

Reagan, Ronald (Mar. 16, 1986). *Address to the Nation on the Situation in Nicaragua*. URL: https://www.reaganlibrary.gov/archives/speech/address-nation-situation-nicaragua.

Requests for User Information – Google Transparency Report (2021). URL: https://transparencyreport.google.com/user-data/overview (visited on 10/12/2021).

RoBERTa (2021). *RoBERTa: An Optimized Method for Pretraining Self-Supervised NLP Systems*. URL: https://ai.facebook.com/blog/roberta-an-optimized-method-for-pretraining-self-supervised-nlp-systems/ (visited on 10/08/2021).

Robertson, Ronald E., Jon Green, et al. (Mar. 19, 2022). "Engagement Outweighs Exposure to Partisan and Unreliable News within Google Search". arXiv: 2201.00074 [cs]. URL: http://arxiv.org/abs/2201.00074 (visited on 03/22/2022).

Robertson, Ronald E., Shan Jiang, et al. (2018). "Auditing Partisan Audience Bias within Google Search". In: *Proceedings of the ACM on Human-Computer Interaction* 2 (CSCW), pp. 1–22.

Robertson, Ronald E., David Lazer, and Christo Wilson (2018). "Auditing the Personalization and Composition of Politically-Related Search Engine Results Pages". In: *Proceedings of the 2018 World Wide Web Conference*, pp. 955–965.

Rodriguez, Pedro L. and Arthur Spirling (Jan. 1, 2022). "Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research". In: *The Journal of Politics* 84.1, pp. 101–115. ISSN: 0022-3816, 1468-2508. DOI: 10.1086/715162. URL: https://www.journals.uchicago.edu/doi/10.1086/715162 (visited on 07/10/2023).

Roehrich, Nico (Aug. 16, 2021). *Apricot Subsea Cable Will Boost Internet Capacity, Speeds in the Asia-Pacific Region*. Facebook Engineering. URL: https://engineering.fb.com/2021/08/15/connectivity/apricot-subsea-cable/ (visited on 10/12/2021).

Rosemain, Matheiu (Oct. 21, 2021). "Facebook Signs Copyright Agreement with Some French Publishers". In: *Reuters*. URL: https://www.reuters.com/technology/facebook-signs-copyright-agreement-with-some-french-publishers-2021-10-21/.

Ruggie, John Gerard (Sept. 2018). "Multinationals as Global Institution: Power, Authority and Relative Autonomy: Multinationals as Global Institution". In: *Regulation & Governance* 12.3, pp. 317–333. ISSN: 17485983. DOI: 10.1111/rego.12154. URL: https://onlinelibrary.wiley.com/doi/10.1111/rego.12154 (visited on 03/16/2023).

Saba, Michel (Dec. 16, 2022). "'Canadians Don't like Being Intimidated': Ottawa Slams Facebook over Threats to Pull News". In: *The Canadian Press*. URL: https://globalnews.ca/news/9355285/facebook-meta-canada-news-content/ (visited on 03/16/2023).

Sandvig, Christian et al. (2014). "Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms". In: *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*. Vol. 22. 64th Annual Meeting of the International Communication Association Seattle, pp. 4349–4357.

Schein, Aaron et al. (Apr. 19, 2021). "Assessing the Effects of Friend-to-Friend Texting on Turnout in the 2018 US Midterm Elections". In: *Proceedings of the Web Conference 2021*. WWW '21: The Web Conference 2021. Ljubljana Slovenia: ACM, pp. 2025–2036. ISBN: 978-1-4503-8312-7. DOI: 10.1145/3442381.3449800. URL: https://dl.acm.org/doi/10.1145/3442381.3449800 (visited on 10/12/2021).

Schelling, Thomas C. (1966). *Arms and Influence*. New Haven, CT: Yale University Press. 312 pp. ISBN: 978-0-300-14337-9.

Schmitt, Carl (2007). *The Concept of the Political*. Expanded ed. Chicago: University of Chicago Press. 126 pp. ISBN: 978-0-226-73892-5.

Scott, James C. (2008). *Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Nachdr. Yale Agrarian Studies. New Haven, Conn.: Yale Univ. Press. 445 pp. ISBN: 978-0-300-07815-2.

Sebastian, Dave (July 9, 2020). "WhatsApp's Business-User Base Grew Tenfold From 2019". In: *The Wall Street Journal*. URL: https://www.wsj.com/articles/whatsapps-business-user-base-grew-tenfold-from-2019-11594298961 (visited on 03/16/2023).

Secretary Rumsfeld Briefs at the Foreign Press Center (Jan. 22, 2003). In collab. with Donald Rumsfeld. URL: https://web.archive.org/web/20140228200949/http://www.defense.gov/transcripts/transcript.aspx?transcriptid=1330.

Shearer, Elisa (Jan. 12, 2021). *More than Eight-in-Ten Americans Get News from Digital Devices*. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices/ (visited on 04/09/2022).

Singh, Kesar (Nov. 1, 1981). "On the Asymptotic Accuracy of Efron's Bootstrap". In: *The Annals of Statistics* 9.6. ISSN: 0090-5364. DOI: 10.1214/aos/1176345636. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-9/issue-6/On-the-Asymptotic-Accuracy-of-Efrons-Bootstrap/10.1214/aos/1176345636.full (visited on 03/29/2023).

Sinha, Satyajit (May 24, 2023). *State of IoT 2023: Number of Connected IoT Devices Growing 16% to 16.7 Billion Globally*. IoT Analytics. URL: https://iot-analytics.com/number-connected-iot-devices/#:~:text=The%20latest%20IoT%20Analytics%20%E2%80%9CState,14.3%20billion%20active%20IoT%20endpoints..

Srivastava, Swati (Nov. 23, 2021). "Algorithmic Governance and the International Politics of Big Tech". In: *Perspectives on Politics*, pp. 1–12. ISSN: 1537-5927, 1541-0986. DOI: 10.1017/S1537592721003145. URL: https://www.cambridge.org/core/product/identifier/S1537592721003145/type/journal_article (visited on 07/03/2023).

Strange, Susan (1988). *States and Markets*. London: Pinter Publishers. 263 pp. ISBN: 978-0-86187-942-7 978-0-86187-992-2.

Sweney, Mark (Oct. 5, 2021). "Facebook Outage Highlights Global Over-Reliance on Its Services". In: *The Guardian*. URL: https://www.theguardian.com/technology/2021/oct/05/facebook-outage-highlights-global-over-reliance-on-its-services (visited on 03/16/2023).

Tang, Min and Narisong Huhe (Nov. 1, 2014). "Alternative Framing: The Effect of the Internet on Political Support in Authoritarian China". In: *International Political Science Review* 35.5, pp. 559–576. ISSN: 0192-5121. DOI: 10.1177/0192512113501971. URL: https://doi.org/10.1177/0192512113501971 (visited on 10/12/2021).

Tannenwald, Nina (1999). "The Nuclear Taboo: The United States and the Normative Basis of Nuclear Non-Use". In: *International Organization* 53.3, pp. 433–468. ISSN: 0020-8183. JSTOR: 2601286. URL: https://www.jstor.org/stable/2601286 (visited on 10/07/2021).

Terman, Rochelle and Joshua Byun (May 2022). "Punishment and Politicization in the International Human Rights Regime". In: *American Political Science Review* 116.2, pp. 385–402. ISSN: 0003-0554, 1537-5943. DOI: 10.1017/S0003055421001167. URL: https://www.cambridge.org/core/product/identifier/S0003055421001167/type/journal_article (visited on 03/29/2023).

The ICRC's Funding and Spending (Apr. 11, 2016). International Committee of the Red Cross. URL: https://www.icrc.org/en/faq/icrcs-funding-and-spending (visited on 10/12/2021).

The United States and France: Allies, Partners, and Friends (June 24, 2021). U.S. Department of State Office of the Spokesperson. URL: https://www.state.gov/the-united-states-and-france-allies-partners-and-friends/.

Tibshirani, Robert, Guenther Walther, and Trevor Hastie (July 1, 2001). "Estimating the Number of Clusters in a Data Set Via the Gap Statistic". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 63.2, pp. 411–423. ISSN: 1369-7412, 1467-9868. DOI: 10.1111/1467-9868.00293. URL: https://academic.oup.com/jrsssb/article/63/2/411/7083348 (visited on 06/27/2023).

Tilley, Aaron (Mar. 8, 2022). "Amazon Stops Accepting New AWS Customers From Russia and Belarus". In: *The Wall Street Journal*. URL: https://www.wsj.com/livecoverage/russia-ukraine-latest-news-2022-03-08/card/amazon-stops-accepting-new-aws-customers-from-russia-and-belarus-Xkq5FmSTc5rZ9pGQu4MS.

Tomz, Michael, Jessica L.P. Weeks, and Keren Yarhi-Milo (2020). "Public Opinion and Decisions About Military Force in Democracies". In: *International Organization* 74.1, pp. 119–143. ISSN: 0020-8183, 1531-5088. DOI: 10.1017/S0020818319000341. URL: https://www.cambridge.org/core/product/identifier/S0020818319000341/type/journal_article (visited on 07/28/2023).

Trefis (Mar. 2, 2020). "How Much In Online Revenue Can Walmart Generate In 2020?" In: *Forbes*. URL: https://www.forbes.com/sites/greatspeculations/2020/03/02/how-much-in-online-revenue-can-walmart-generate-in-2020/?sh=3937fdc12e26.

Trielli, Daniel and Nicholas Diakopoulos (2019). "Search as News Curator: The Role of Google in Shaping Attention to News Information". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–15.

Trielli, Daniel and Nicholas Diakopoulos (Jan. 2, 2022). "Partisan Search Behavior and Google Results in the 2018 U.S. Midterm Elections". In: *Information, Communication & Society* 25.1, pp. 145–161. ISSN: 1369-118X, 1468-4462. DOI: 10.1080/1369118X.2020.1764605. URL: https://www.tandfonline.com/doi/full/10.1080/1369118X.2020.1764605 (visited on 08/10/2023).

Tripodi, Francesca (2018). *Searching for Alternative Facts: Analyzing Scriptural Inference in Conservative News Practices*. Data & Society.

Trugeur, Felix (2019). "Seeing Like Big Tech: Security Assemblages, Technology, and the Future of State Bureaucracy". In: *Data Politics: Worlds, Subjects, Rights*. Routledge Studies in International Political Sociology. London ; New York: Routledge, Taylor & Francis Group. ISBN: 978-1-138-05325-0 978-1-138-05326-7.

Turowski, Jacob (Feb. 1, 2023). *Why Claims of Bias in Our Content Review Process Are Wrong*. Meta. URL: https://about.fb.com/news/2023/02/why-claims-of-bias-in-our-content-review-process-are-wrong/.

U.S. Relations With Israel (Jan. 20, 2021). U.S. Department of State Bureau of Near Eastern Affairs. URL: https://www.state.gov/u-s-relations-with-israel/ (visited on 03/28/2023).

Understand & Manage Your Location When You Search on Google - Computer - Google Search Help (2022). URL: https://support.google.com/websearch/answer/179386?hl=en-IR (visited on 04/08/2022).

Van Couvering, Elizabeth (2007). "Is Relevance Relevant? Market, Science, and War: Discourses of Search Engine Quality". In: *Journal of Computer-Mediated Communication* 12.3, pp. 866–887. ISSN: 1083-6101. DOI: 10.1111/j.1083-6101.2007.00354.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1083-6101.2007.00354.x (visited on 04/09/2022).

Vaughan, Liwen and Mike Thelwall (July 2004). "Search Engine Coverage Bias: Evidence and Possible Causes". In: *Information Processing & Management* 40.4, pp. 693–707. ISSN: 03064573. DOI: 10.1016/S0306-4573(03)00063-3. URL: https://linkinghub.elsevier.com/retrieve/pii/S0306457303000633 (visited on 04/09/2022).

Vaughan, Liwen and Yanjun Zhang (Apr. 2007). "Equal Representation by Search Engines? A Comparison of Websites across Countries and Domains". In: *Journal of Computer-Mediated Communication* 12.3, pp. 888–909. ISSN: 10836101, 10836101. DOI: 10.1111/j.1083-6101.2007.00355.x. URL: https://academic.oup.com/jcmc/article/12/3/888-909/4583004 (visited on 08/10/2023).

Verma, Nitin, Kenneth R. Fleischmann, and Kolina S. Koltai (Jan. 2017). "Human Values and Trust in Scientific Journals, the Mainstream Media and Fake News". In: *Pro-*

ceedings of the Association for Information Science and Technology 54.1, pp. 426–435. ISSN: 2373-9231, 2373-9231. DOI: 10.1002/pra2.2017.14505401046. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/pra2.2017.14505401046 (visited on 02/16/2021).

Vidino, Lorenzo and Hughes Seamus (2015). "ISIS in America: From Retweets to Raqqa". In: *George Washington University: Program on Extremism*.

Voeten, Erik (2021). *Ideology and International Institutions*. Princeton, New Jersey: Princeton University Press. 242 pp. ISBN: 978-0-691-20731-5 978-0-691-20732-2.

Vosoughi, Soroush, Deb Roy, and Sinan Aral (Mar. 9, 2018). "The Spread of True and False News Online". In: *Science* 359.6380, pp. 1146–1151. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aap9559. URL: https://www.sciencemag.org/lookup/doi/10.1126/science.aap9559 (visited on 02/16/2021).

Waller, Isaac and Ashton Anderson (Dec. 9, 2021). "Quantifying Social Organization and Political Polarization in Online Platforms". In: *Nature* 600.7888, pp. 264–268. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-021-04167-x. URL: https://www.nature.com/articles/s41586-021-04167-x (visited on 03/29/2023).

Waltz, Kenneth Neal (1979). *Theory of International Politics*. Reiss. Long Grove, Ill: Waveland Press. 251 pp. ISBN: 978-1-57766-670-7.

War Project, Correlates of (2017). *State System Membership List, V2016.* Online.

Warner, Benjamin R. (Aug. 17, 2010). "Segmenting the Electorate: The Effects of Exposure to Political Extremism Online". In: *Communication Studies* 61.4, pp. 430–444. ISSN: 1051-0974, 1745-1035. DOI: 10.1080/10510974.2010.497069. URL: http://www.tandfonline.com/doi/abs/10.1080/10510974.2010.497069 (visited on 08/03/2023).

Webber, William, Alistair Moffat, and Justin Zobel (Nov. 2010). "A Similarity Measure for Indefinite Rankings". In: *ACM Transactions on Information Systems* 28.4, pp. 1–38. ISSN: 1046-8188, 1558-2868. DOI: 10.1145/1852102.1852106. URL: https://dl.acm.org/doi/10.1145/1852102.1852106 (visited on 11/14/2022).

Weber, Ingmar and Alejandro Jaimes (2011). "Who Uses Web Search for What: And How". In: *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, pp. 15–24.

Weber, Max, Guenther Roth, and Claus Wittich (1978). *Economy and society: an outline of interpretive sociology*. Berkeley: University of California Press. 2 pp. ISBN: 978-0-520-02824-1 978-0-520-03500-3.

Weinstein, Jeremy M. (2007). *Inside Rebellion: The Politics of Insurgent Violence*. Cambridge Studies in Comparative Politics. Cambridge ; New York: Cambridge University Press. 402 pp. ISBN: 978-0-521-86077-2 978-0-521-67797-4.

Wendt, Alexander (June 1994). "Collective Identity Formation and the International State". In: *American Political Science Review* 88.2, pp. 384–396. ISSN: 0003-0554, 1537-5943. DOI: 10.2307/2944711. URL: https://www.cambridge.org/core/product/identifier/S0003055400092807/type/journal_article (visited on 03/29/2023).

— (1999). *Social Theory of International Politics*. ISBN: 978-0-511-00417-9 978-0-511-03775-7 978-0-511-05299-6 978-0-511-61218-3. URL: https://doi.org/10.1017/CBO9780511612183 (visited on 02/16/2021).

White, Aoife and Stephanie Bodini (Sept. 28, 2021). "Google Tells Judges It's So Popular It's Bing's Top Search Term". In: *Bloomberg News*. URL: https://www.bloomberg.com/news/articles/2021-09-28/google-tells-judges-it-s-so-popular-it-s-bing-s-top-search-term (visited on 10/01/2021).

Xin, Xin (Dec. 1, 2009). "Xinhua News Agency in Africa". In: *Journal of African Media Studies* 1.3, pp. 363–377. ISSN: 2040-199X, 1751-7974. DOI: 10.1386/jams.1.3.363/1. URL: https://intellectdiscover.com/content/journals/10.1386/jams.1.3.363/1 (visited on 08/01/2023).

Xing, Xinyu et al. (2014). "Exposing Inconsistent Web Search Results with Bobble". In: *International Conference on Passive and Active Network Measurement*, pp. 131–140.

Yearwood, Maurice H. et al. (Dec. 2015). "On Wealth and the Diversity of Friendships: High Social Class People around the World Have Fewer International Friends". In: *Personality and Individual Differences* 87, pp. 224–229. ISSN: 01918869. DOI: 10.1016/j.paid.2015.07.040. URL: https://linkinghub.elsevier.com/retrieve/pii/S0191886915004973 (visited on 07/03/2023).

Zarpli, Omer and Huseyin Zengin (Jan. 2022). "Shame, Endorse, or Remain Silent?: State Response to Human Rights Violations in Other Countries". In: *Research & Politics* 9.1, p. 205316802110703. ISSN: 2053-1680, 2053-1680. DOI: 10.1177/20531680211070344. URL: http://journals.sagepub.com/doi/10.1177/20531680211070344 (visited on 03/29/2023).

Ziegler, Charles E. (Sept. 2018). "International Dimensions of Electoral Processes: Russia, the USA, and the 2016 Elections". In: *International Politics* 55.5, pp. 557–574. ISSN: 1384-5748, 1740-3898. DOI: 10.1057/s41311-017-0113-1. URL: http://link.springer.com/10.1057/s41311-017-0113-1 (visited on 02/16/2021).

Zuboff, Shoshana (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. First edition. New York: PublicAffairs. 691 pp. ISBN: 978-1-61039-569-4.

Zuckerberg, Mark (Oct. 6, 2021). *I Wanted to Share a Note I Wrote to Everyone at Our Company.* URL: `https://www.facebook.com/zuck` (visited on 10/08/2021).

Appendices

Appendix A

Appendix to Chapter 3

A.1 Search Locations and Parameters

Table A.1: Search Locations and Parameters

Country	SerpAPI Location	Google Domain	Language
Afghanistan	Kabul, Kabul, Afghanistan	google.com.af	Persian
Albania	Tirana County, Albania	google.al	Albanian
Algeria	Algiers Province, Algeria	google.dz	Arabic
Andorra	Andorra	google.ad	Catalan
Angola	Luanda, Luanda Province, Angola	google.co.ao	Portuguese
Antigua and Barbuda	Saint John's, Antigua and Barbuda	google.com.ag	English
Argentina	Buenos Aires, Buenos Aires, Argentina	google.com.ar	Spanish
Armenia	Yerevan, Armenia	google.am	Armenian
Australia	Canberra, Australian Capital Territory, Australia	google.com.au	English
Austria	Vienna, Vienna, Austria	google.at	German
Azerbaijan	Baku, Azerbaijan	google.az	Azerbaijani
Bahamas	The Bahamas	google.bs	English
Bahrain	Manama, Capital Governorate, Bahrain	google.com.bh	Arabic
Bangladesh	Dhaka District, Dhaka Division, Bangladesh	google.com.bd	Bangla
Barbados	Saint Michael, Barbados	google.com	English
Belarus	Minsk, Minsk Region, Belarus	google.by	Russian
Belgium	Brussels, Brussels, Belgium	google.be	Dutch
Belize	Cayo District, Belize	google.com.bz	English

Table A.1: Search Locations and Parameters (continued)

Country	SerpAPI Location	Google Domain	Language
Benin	Porto Novo, Oueme Department, Benin	google.bj	French
Bolivia	La Paz Department, Bolivia	google.com.bo	Spanish
Bosnia and Herzegovina	Bosnia and Herzegovina	google.ba	Bosnian
Brazil	Brasilia, Federal District, Brazil	google.com.br	Portuguese
Brunei	Bandar Seri Begawan, Brunei-Muara District, Brunei	google.com.bn	Malay
Bulgaria	Sofia, Sofia-Capital, Bulgaria	google.bg	Bulgarian
Cambodia	Phnom Penh, Phnom Penh, Cambodia	google.com.kh	Khmer
Cameroon	Centre, Cameroon	google.cm	French
Canada	Ottawa, Ontario, Canada	google.ca	English
Cape Verde	Praia, Cape Verde	google.cv	Portuguese
Chad	N'Djamena, Chad	google.td	French
Chile	Santiago Metropolitan Region, Chile	google.cl	Spanish
Colombia	Bogota, Bogota, Colombia	google.com.co	Spanish
The Democratic Republic of the Congo	Kinshasa, Democratic Republic of the Congo	google.cd	French
Costa Rica	San Jose, San Jose Province, Costa Rica	google.co.cr	Spanish
Cote D'ivoire	Bouake, Vallee du Bandama District, Cote d'Ivoire	google.ci	French
Croatia	Zagreb, City of Zagreb, Croatia	google.hr	Croatian
Cyprus	Nicosia, Nicosia, Cyprus	google.com.cy	Greek
Czech Republic	Prague, Prague, Czechia	google.cz	Czech
Denmark	Copenhagen, Capital Region of Denmark, Denmark	google.dk	Danish
Djibouti	Djibouti, Djibouti	google.dj	Somali
Dominica	Roseau, Dominica	google.dm	English
Dominican Republic	Santo Domingo Province, Dominican Republic	google.com.do	Spanish
Ecuador	Quito, Pichincha, Ecuador	google.com.ec	Spanish

Table A.1: Search Locations and Parameters (continued)

Country	SerpAPI Location	Google Domain	Language
Egypt	Cairo, Cairo Governorate, Egypt	google.com.eg	Arabic
El Salvador	San Salvador, San Salvador Department, El Salvador	google.com.sv	Spanish
Equatorial Guinea	Equatorial Guinea	google.com	Spanish
Estonia	Tallinn, Harju County, Estonia	google.ee	Estonian
Ethiopia	Addis Ababa, Ethiopia	google.com.et	Amharic
Finland	Helsinki, Helsinki, Finland	google.fi	Finnish
France	Paris, Paris, Ile-de-France, France	google.fr	French
Georgia	Tbilisi, Tbilisi, Georgia	google.ge	Georgian
Germany	Berlin, Berlin, Germany	google.de	German
Greece	Athens, Athens, Attica, Greece	google.gr	Greek
Grenada	Saint George's, Grenada	google.com	English
Guatemala	Guatemala City, Guatemala Department, Guatemala	google.com.gt	Spanish
Guinea	Conakry, Guinea	google.com	French
Guinea Bissau	Guinea-Bissau	google.com	Portuguese
Guyana	Georgetown, Guyana	google.gy	English
Haiti	Port-au-Prince, Ouest Department, Haiti	google.ht	French
Honduras	Francisco Morazan Department, Honduras	google.hn	Spanish
Hungary	Budapest, Budapest, Hungary	google.hu	Hungarian
Iceland	Reykjavik, Capital Region, Iceland	google.is	Icelandic
India	New Delhi, Delhi, India	google.co.in	Hindi
Indonesia	Jakarta, Jakarta, Indonesia	google.co.id	Indonesian
Iraq	Baghdad, Baghdad Governorate, Iraq	google.iq	Arabic
Ireland	Dublin, County Dublin, Ireland	google.ie	English
Israel	Jerusalem, Jerusalem District, Israel	google.co.il	Hebrew
Italy	Rome, Lazio, Italy	google.it	Italian

Table A.1: Search Locations and Parameters (continued)

Country	SerpAPI Location	Google Domain	Language
Japan	Tokyo, Japan	google.co.jp	Japanese
Jordan	Amman Governorate, Jordan	google.jo	Arabic
Kazakhstan	Astana, Kazakhstan	google.kz	Russian
Kenya	Nairobi, Nairobi County, Kenya	google.co.ke	Swahili
Republic of Korea	Seoul, Seoul, South Korea	google.co.kr	Korean
Kuwait	Al Asimah Governate, Kuwait	google.com.kw	Arabic
Kyrgyzstan	Bishkek, Chuy Province, Kyrgyzstan	google.kg	Kyrgyz
Laos	Vientiane, Vientiane Prefecture, Laos	google.la	Lao
Latvia	Riga, Riga, Latvia	google.lv	Latvian
Lebanon	Beirut, Beirut Governorate, Lebanon	google.com.lb	Arabic
Liberia	Monrovia, Liberia	google.com	English
Libya	Tripoli, Tripoli District, Libya	google.com.ly	Arabic
Liechtenstein	Liechtenstein	google.li	German
Lithuania	Vilnius, Vilnius County, Lithuania	google.lt	Lithuanian
Luxembourg	Luxembourg City, Luxembourg	google.lu	Luxembourgish
Macedonia	Macedonia (FYROM)	google.mk	Macedonian
Madagascar	Antananarivo, Antananarivo Province, Madagascar	google.mg	Malagasy
Malaysia	Kuala Lumpur, Federal Territory of Kuala Lumpur, Malaysia	google.com.my	Malay
Malta	Valletta, Malta	google.com.mt	Maltese
Mauritania	Nouakchott, Mauritania	google.com	Arabic
Mexico	Mexico City, Mexico City, Mexico	google.com.mx	Spanish
Moldova	Chisinau, Moldova	google.md	Romanian
Monaco	Monaco	google.com	French
Mongolia	Ulaanbaatar, Mongolia	google.mn	Mongolian
Morocco	Rabat, Rabat-Sale-Kenitra, Morocco	google.co.ma	Arabic

Table A.1: Search Locations and Parameters (continued)

Country	SerpAPI Location	Google Domain	Language
Mozambique	Maputo, Mozambique	google.co.mz	Portuguese
Myanmar	Naypyitaw, Naypyitaw Union Territory, Myanmar (Burma)	google.com.mm	Burmese
Nepal	Kathmandu, Central Development Region, Nepal	google.com.np	Nepali
Netherlands	Amsterdam, North Holland, Netherlands	google.nl	Dutch
New Zealand	Wellington, Wellington, New Zealand	google.co.nz	English
Nicaragua	Managua, Managua Department, Nicaragua	google.com.ni	Spanish
Niger	Niamey, Niamey Urban Community, Niger	google.ne	Hausa
Nigeria	Abuja, Federal Capital Territory, Nigeria	google.com.ng	Hausa
Norway	Oslo, Oslo, Norway	google.no	Norwegian
Oman	Muscat Governorate, Oman	google.com.om	Arabic
Pakistan	Islamabad, Islamabad Capital Territory, Pakistan	google.com.pk	Punjabi
Palestine	Palestine	google.ps	Arabic
Panama	Panama, Panama	google.com.pa	Spanish
Peru	Lima, Lima Province, Peru	google.com.pe	Spanish
Philippines	Manila, Metro Manila, Philippines	google.com.ph	Filipino
Poland	Warsaw, Masovian Voivodeship, Poland	google.pl	Polish
Portugal	Lisbon, Lisbon, Portugal	google.pt	Portuguese
Qatar	Doha, Doha, Qatar	google.com.qa	Arabic
Romania	Bucharest, Bucharest, Romania	google.ro	Romanian
Russia	Moscow, Moscow, Russia	google.ru	Russian
Rwanda	Kigali, Kigali City, Rwanda	google.rw	Kinyarwanda
Samoa	Apia, Samoa	google.ws	Samoan
San Marino	San Marino	google.sm	Italian
Sao Tome and Principe	Sao Tome, Sao Tome and Principe	google.com	Portuguese
Saudi Arabia	Riyadh, Riyadh Province, Saudi Arabia	google.com.sa	Arabic

Table A.1: Search Locations and Parameters (continued)

Country	SerpAPI Location	Google Domain	Language
Serbia	Serbia	google.rs	Serbian
Singapore	Singapore	google.com.sg	English
Slovakia	Bratislava, Bratislava Region, Slovakia	google.sk	Slovak
Slovenia	Ljubljana, Ljubljana, Slovenia	google.si	Slovenian
Somalia	Mogadishu, Somalia	google.so	Somali
South Africa	Pretoria, Gauteng, South Africa	google.co.za	Zulu
Spain	Madrid, Community of Madrid, Spain	google.es	Spanish
Sri Lanka	Colombo, Western Province, Sri Lanka	google.lk	Sinhala
Sudan	Khartoum, Sudan	google.com	Arabic
Sweden	Stockholm, Stockholm County, Sweden	google.se	Swedish
Switzerland	Bern, Canton of Bern, Switzerland	google.ch	German
Taiwan	Taipei City, Taiwan	google.com.tw	Chinese (Traditional)
Tajikistan	Dushanbe, Districts of Republican Subordination, Tajikistan	google.com.tj	Tajik
Tanzania	Dodoma, Dodoma Region, Tanzania	google.co.tz	Swahili
Thailand	Bangkok, Bangkok, Thailand	google.co.th	Thai
Togo	Lome, Maritime Region, Togo	google.tg	French
Tunisia	Tunis, Tunisia	google.tn	French
Turkey	Ankara, Ankara, Turkey	google.com.tr	Turkish
Turkmenistan	Ashgabat, Turkmenistan	google.tm	Turkmen
Uganda	Kampala, Central Region, Uganda	google.co.ug	English
Ukraine	Kyiv, Kyiv city, Ukraine	google.com.ua	Ukrainian
United Arab Emirates	Abu Dhabi, Abu Dhabi, United Arab Emirates	google.ae	English
United Kingdom	London, England, United Kingdom	google.co.uk	English

Table A.1: Search Locations and Parameters (continued)

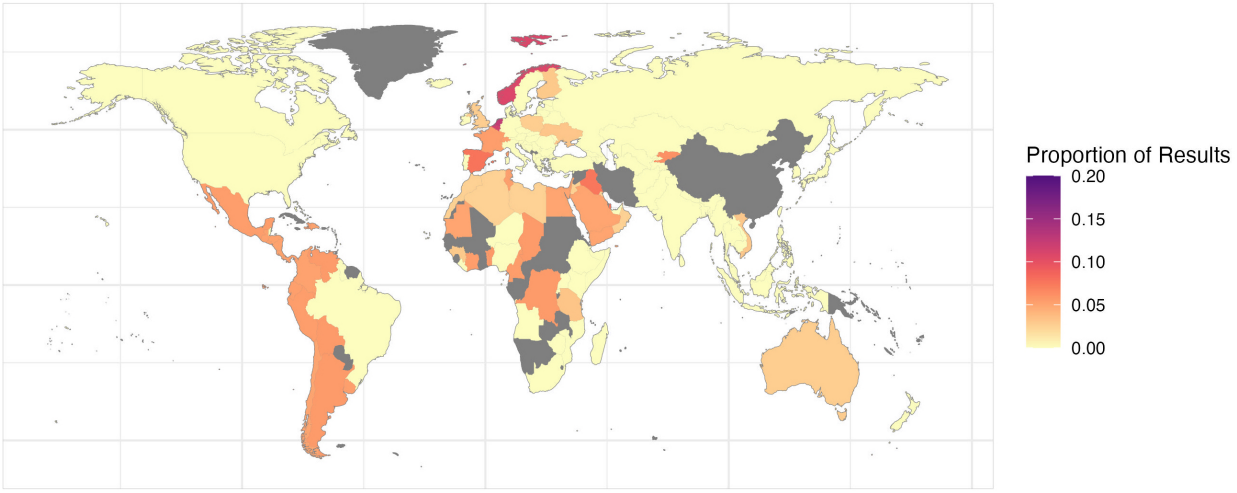
Country	SerpAPI Location	Google Domain	Language
United States	Washington, District of Columbia, United States	google.com	English
Uruguay	Montevideo, Uruguay	google.com.uy	Spanish
Uzbekistan	Tashkent, Uzbekistan	google.co.uz	Uzbek
Venezuela	Caracas, Capital District, Venezuela	google.co.ve	Spanish
Viet Nam	Hanoi, Hanoi, Vietnam	google.com.vn	Vietnamese
Yemen	Sana'a, Capital Municipality, Yemen	google.com	Arabic
Zimbabwe	Harare, Harare Province, Zimbabwe	google.co.zw	Shona

A.2 Additional Analyses

A.2.1 Ideological Content

In Figure A.1, we assess the reach of transnational advocacy organizations for the search queries "Uyghurs genocide" and "Uyghurs terrorism." Focusing on the same set of publishers as Figure 3.1 in the main text, we find that the overall presence of results from transnational advocacy organizations is lower compared to the human rights queries. However, we still observe a consistent presence of results from transnational advocacy organizations in countries across particular regions and language groups. Countries with Spanish, Arabic, and French-speaking populations appear to consistently exhibit some prevalence of search results from these organizations. This finding further highlights the disparate representation of transnational advocacy organizations in localized search results.

Figure A.1: Prevalence of Transnational Advocacy Organizations in Uyghur Searches



A.2.2 Alternative Network Diagrams

Below, we include alternative versions of the network figures presented in the main analysis. Figure A.2 replicates Figure 3.4 from the main text, instead using RBO to assess the similarity in publishers between the SERPs of two countries. Figures A.3 and A.4 construct the network using the Google Image results and the MCS and CS-RBO metrics, respectively.

Figure A.2: Network of Source Similarity (Source RBO)

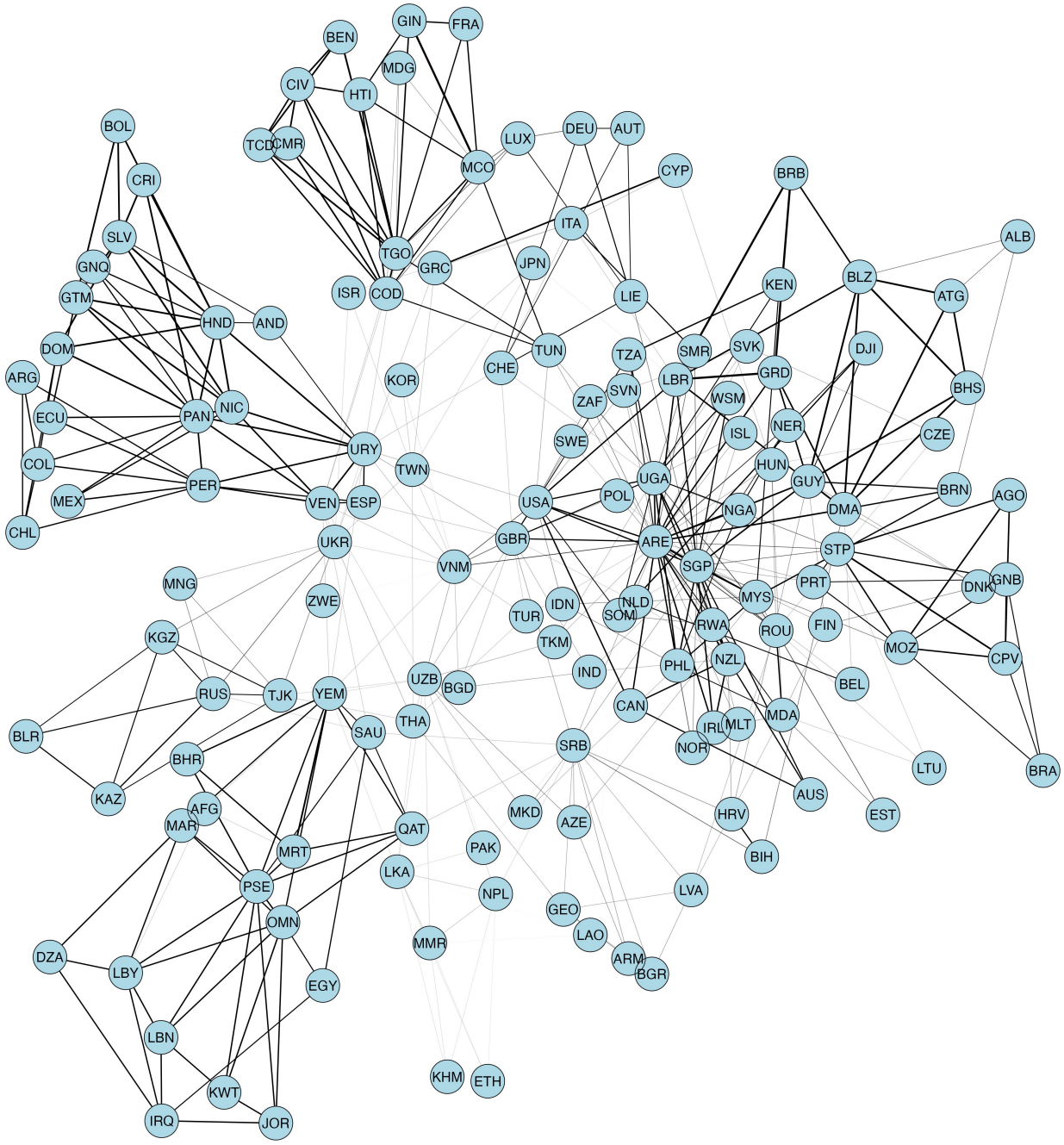
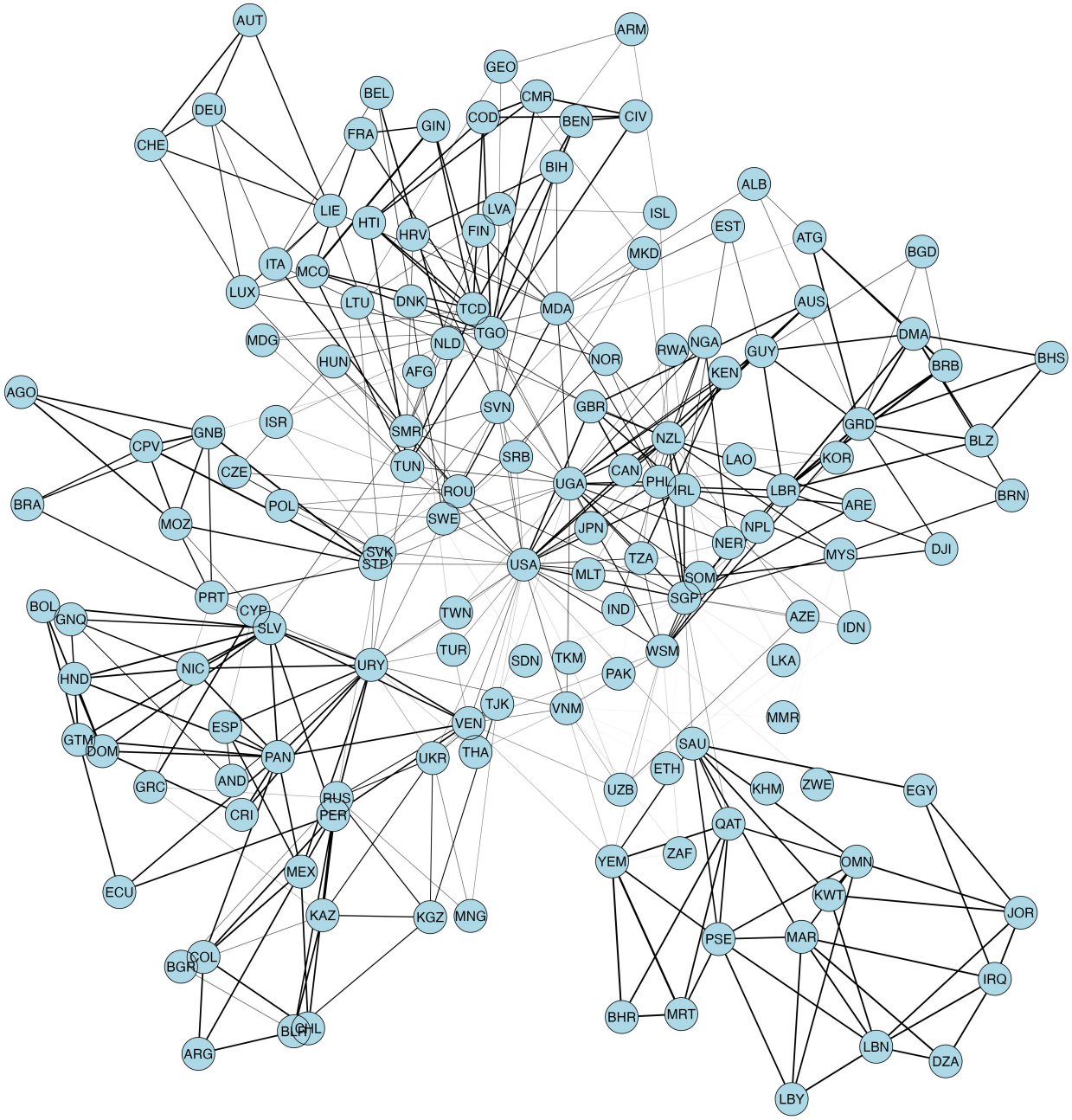


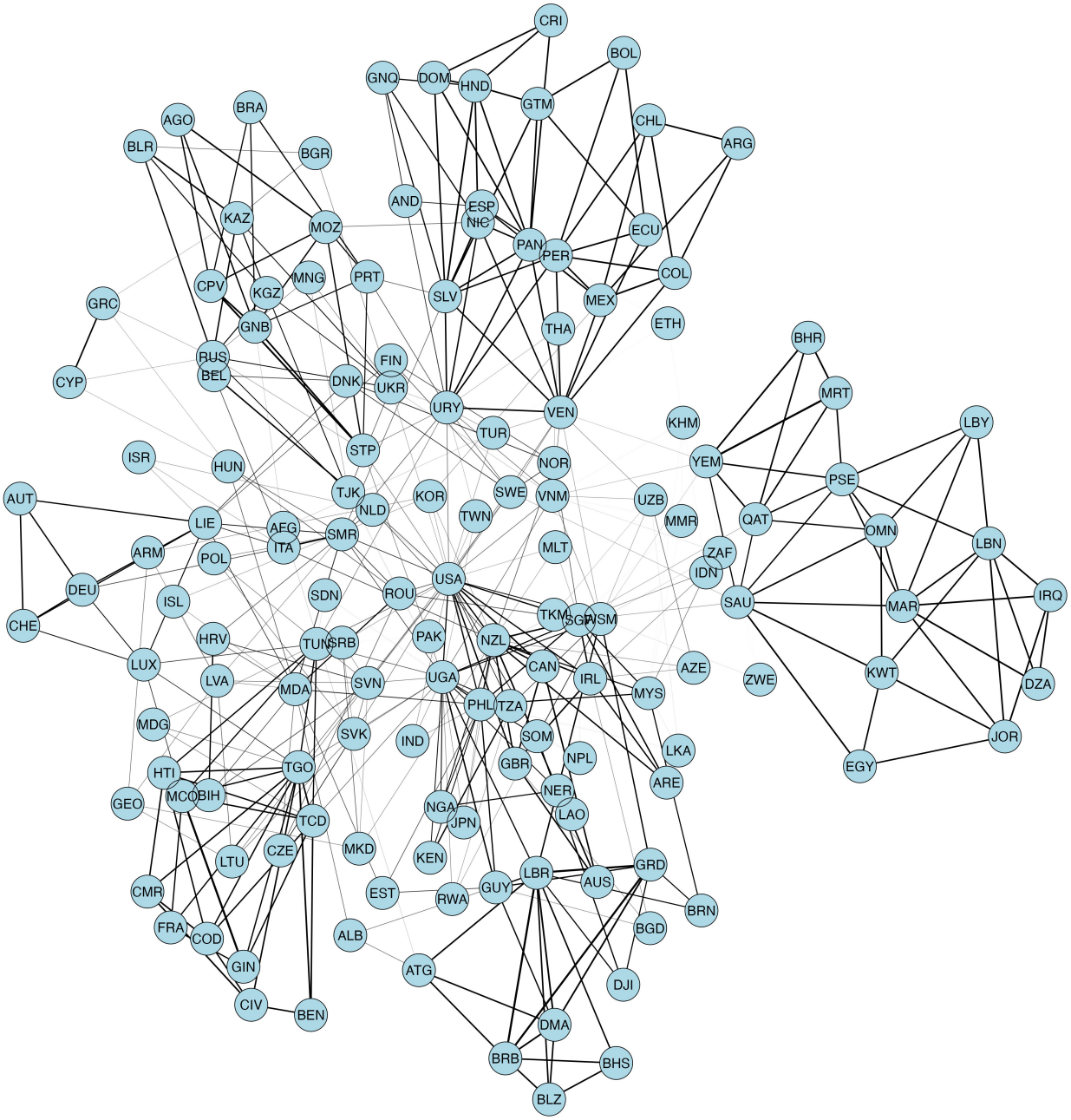
Figure A.3: Network of Image Similarity (MCS)



A.2.3 Evaluating Clustering Quality

In the main text, we evaluate the clustering behavior of countries based upon the publishers present in their SERPs related to the war in Ukraine. To better assess any changes

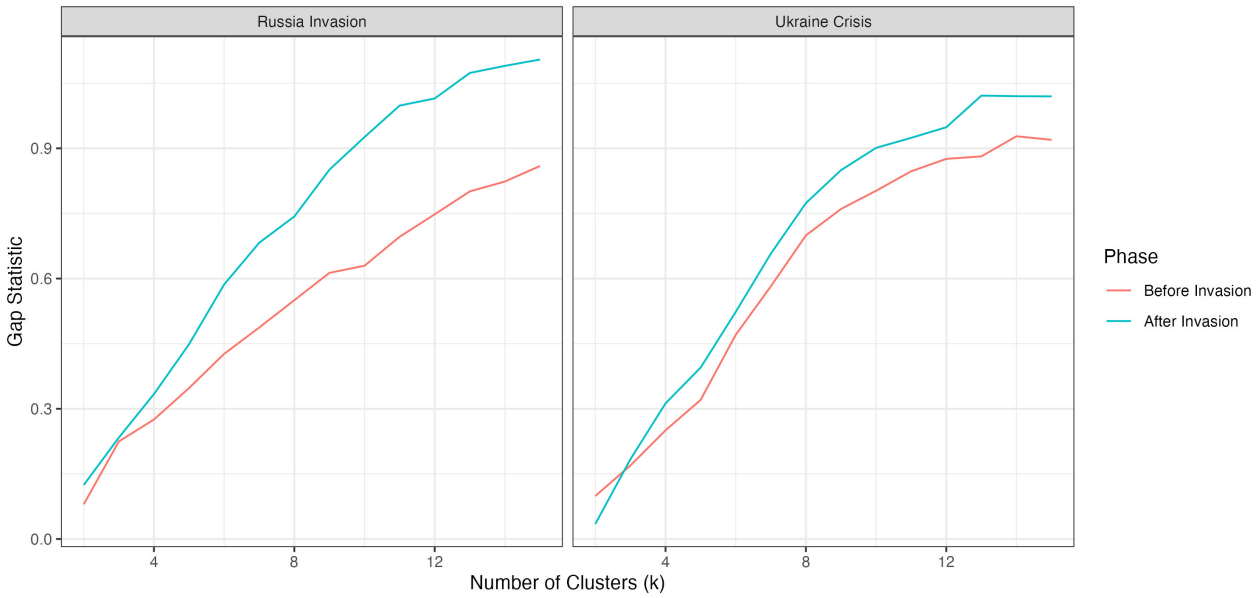
Figure A.4: Network of Image Similarity (CS-RBO)



in clustering between our two rounds of searches, we conduct two exercises to assess the quality of clustering in the results. First, we run k-means clustering on each panel in Figures 3.5 and 3.6 and evaluate the gap statistic of the clusters. The gap statistic serves as an

estimation of the optimal number of clusters in a data set by comparing the within-cluster dispersion to that expected from a null reference distribution.[1] We repeat this procedure for all values of k between 2 and 15. Figure A.5 presents the results. While we do not claim a specific number of optimal clusters for each panel, we observe that for nearly all choices of k, the post-invasion data displays better-defined clusters. This is evident in the higher gap statistic, indicating improved clustering, for all options of k in the "Russia invasion" searches, and for all options greater than for 2 in the "Ukraine crisis" searches.

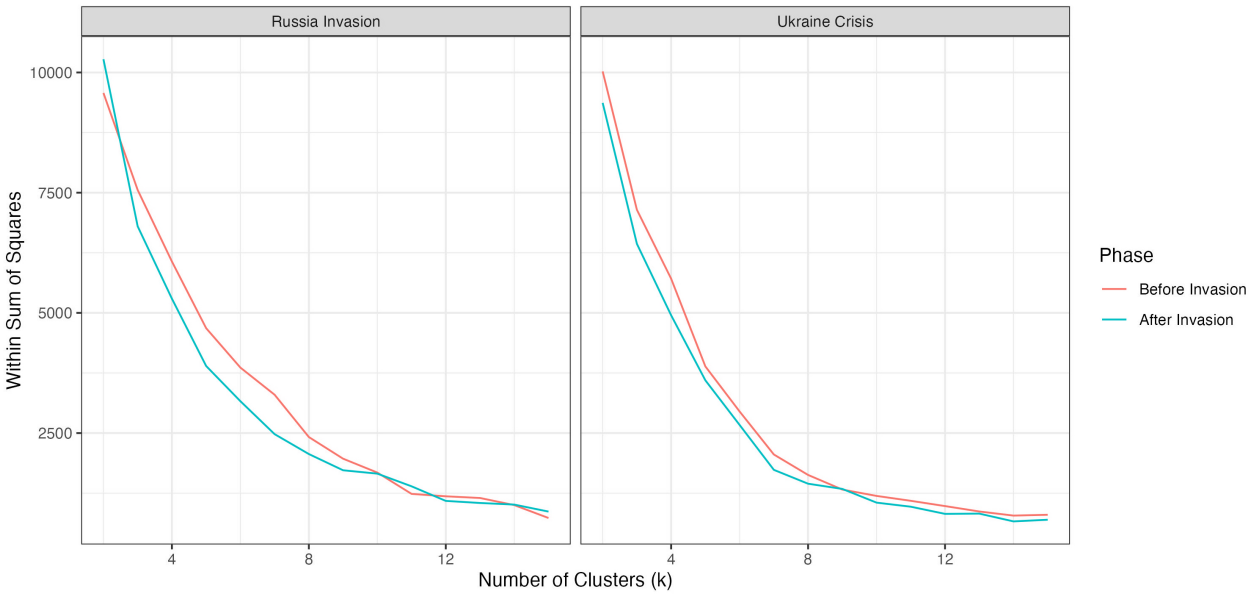
Figure A.5: Gap Statistic for Different Numbers of Clusters



We find a similar result in our second exercise, where we evaluate the within-cluster sum of squares (WSS) across different choices of k. We present the results in Figure A.6. For both search queries, the majority of k choices exhibit stronger clustering post-invasion, indicated by lower values for WSS. The WSS value is lower after the invasion for all values of k other than 9 in the "Ukraine crisis" searches, and all values of k between 3 and 10 for the "Russia invasion" searches.

[1] Tibshirani, Walther, and Hastie 2001.

Figure A.6: WSS for Different Numbers of Clusters



A.3 Alternative Regression Models

Below, we present two additional regression models with alternative dependent variable specifications. In Model 1, we use RBO on the publisher rankings as the dependent variable, taking the mean RBO value across all worldview SERPs for the dyad. Model 2 takes the mean CS-RBO measure for the image results across all queries for the dyad as the dependent variable. We include the same independent variables in these models as those in the main text. Table A.2 presents the results.

Table A.2: Correlates of SERP Variation with Alternative Dependent Variables

	Dependent Variable	
	Source RBO	Image CS-RBO
	(1)	(2)
Common Language	0.190***	0.048***
	(0.004)	(0.001)
Total Bilateral Trade Flows (log)	0.0002	0.001***
	(0.001)	(0.0002)
Colonial History	0.042***	0.010***
	(0.011)	(0.003)
Social Connectedness	0.008***	0.003***
	(0.001)	(0.0003)
Same Region	0.030***	0.010***
	(0.004)	(0.001)
GDP Difference (log)	−0.002**	0.0004
	(0.001)	(0.0002)
Constant	0.286***	0.080***
	(0.020)	(0.007)
Observations	7,022	7,430
R^2	0.362	0.298
Note:	*p<0.1; **p<0.05; ***p<0.01	

The effect of *Common Language* is positive and statistically significant at conventional levels in both models, providing further support for the hypothesis that countries sharing the same language tend to have more similar SERPs. As in the main text, this finding holds for both similarity in the content publishers as well as the image content. *Total Bilateral Trade Flows* only reaches conventional levels of statistical significance in Model 2, which again suggests that trade relationships are associated with differences in SERP content but not in the publishers that populate the SERPs. The positive coefficient of *Colonial History* in both models indicates that countries with shared colonial ties exhibit greater similarity in their SERPs. The coefficients on *Social Connectedness* and *Same Region* are of the same direction as the main text and reach the same levels of statistical significance. Finally, the coefficient on *GDP Difference* only reaches conventional levels of significance in Model 1, providing limited evidence of a relationship between GDP similarity and SERP similarity.

A.4 Image Similarity Metrics

This appendix describes the maximum cosine similarity (MCS) and cosine similarity rank biased overlap (CS-RBO) measurements that we use to assess the similarity of two SERPs. MCS provides a measure of similarity by comparing pairs of image embedding vectors from SERPs based on cosine similarity. Let *MCS(*$SERP_1$*,* $SERP_2$*)* represent the MCS function between the image results of two SERPs. The function takes two lists as inputs, $SERP_1$ and $SERP_2$, which each contain a list of the embeddings of the Google Image results of a SERP.

$$\text{MCS}(\text{SERP}_1, \text{SERP}_2) = \sum_{d=1}^{l} \max\left(\text{CS-Vector}\right)$$

Where:

- $SERP_1$ and $SERP_2$ are the input lists of embeddings.

- l is the number of iterations, which is the minimum of the lengths of $SERP_1$ and $SERP_2$.
- *CS-Vector* is a vector storing the cosine similarity values for all combinations of embedding vectors between $SERP_1$ and $SERP_2$.

During each iteration, the function finds the maximum value from *CS-Vector*, representing the highest cosine similarity value among all the embedding pairs. After storing this value, the function discards the pair of embedding vectors that contributed to the selected value and then repeats this process for l iterations. This removal process guarantees that the previously selected images and their associated embeddings do not contribute to the selected cosine similarity values in subsequent iterations. The final result is the sum of all the maximum cosine similarity values obtained during the iterations.

CS-RBO provides an alternative measure of similarity that incorporates both the MCS of the Google Image results as well as a ranking bias. This metric is a variation of Rank Biased Overlap (RBO), modified to incorporate the cosine similarity of each pair of list entries instead of discrete entry intersections. The CS-RBO computation involves recursively calculating a weighted sum of cosine similarities at different list depths, with a weighting parameter that reflects the importance of top rankings. This process enables us to capture the overall similarity between the two lists while taking into account the rank order of the images.

Let *CS-RBO($SERP_1$, $SERP_2$, p)* represent the CS-RBO function between the image results of two SERPs. The function takes the same two lists as *MCS($SERP_1$, $SERP_2$)* as inputs, as well as a weighting parameter p.

$$\text{CS-RBO}(\text{SERP}_1, \text{SERP}_2, p) = \frac{1-p}{p} \left(\sum_{d=1}^{k} p^d \cdot \frac{\text{MCS}_d}{d} \right) + \frac{\text{MCS}_k}{k} \cdot p^k$$

Where:

- p is the parameter that determines the rank bias, with a value between 0 and 1. A value closer to 0 indicates greater top-weighting of the results, while a value of 1 removes rank weighting. The calculations in this chapter were performed with a p of 0.9.
- k is the maximum length between $SERP_1$ and $SERP_2$.
- MCS_d is the maximum cosine similarity of $SERP_1$ and $SERP_2$ up to list depth d.

The CS-RBO function combines the maximum cosine similarity term (weighted by the ratio of MCS_k to k and p^k) with a summation term. The summation considers the maximum cosine similarities at different ranks (weighted by p^d) and accumulates them for ranks from 1 to k. The $\frac{1-p}{p}$ factor adjusts the weighting to account for the rank bias.

Appendix B

Appendix to Chapter 4

B.1 Alternative Models

This appendix presents alternative models based on projections created from a reduced set of antonym pairs. For the friendship dimension, I use a single set of antonyms, *friend–enemy*; for the collective identity dimension, I use *we–they* as the only set of antonyms. However, it is worth noting that Kozlowski et al. (2019) find that dimensions constructed on a single antonym pair fare relatively poorly when validated against ground truth survey data.[1]

Upon reproducing the regression tables from the main text (Table 4.2 and Table 4.3) with projections based on the reduced antonym sets, I find that the results are largely robust. For the friendship models in Table B.1, the direction of the effect of *MID (Last 20 Years)*, *Total Trade Flows (Log)*, and *Alliance* are the same as the main text across all models. There is a notable result in *Common Language*, which exhibits a statistically significant positive effect not observed in the main analysis. While this would suggest that states with different languages are more likely to be perceived as enemies, I urge readers to exercise caution in interpreting this result, given the known instability associated with dimensions built on a single antonym pair. The results for *Distance*, *POLITY Difference*, and *CINC (Scaled)* are largely similar to those presented the main text.

There is more variability between the results presented in Table B.2 and the results from

[1] Kozlowski, Taddy, and J. A. Evans 2019.

the main text for the collective identity dimension. Again, I approach these findings with caution due to the instability associated with these projections. Notably, the coefficients on *Common Language* and *Distance* are most similar to the results presented in the main text.

Table B.1: Correlates of Friendship Projections with Reduced Antonym Set

	Friendship Percentile Rank		
	(1)	(2)	(3)
MID (Last 20 Years)	-16.168^{***}	-16.879^{***}	-8.381^{***}
	(1.011)	(0.988)	(0.970)
Total Trade Flows (Log)	0.144	0.214^{*}	0.918^{***}
	(0.111)	(0.124)	(0.132)
Alliance	0.305	-1.833^{*}	2.781^{**}
	(0.897)	(0.949)	(1.279)
Common Language	3.218^{***}	3.550^{***}	
	(0.783)	(0.763)	
Distance (Log)	-8.593^{***}	-9.516^{***}	
	(0.779)	(0.785)	
POLITY Difference	-0.516^{***}	-0.603^{***}	-0.290^{***}
	(0.054)	(0.055)	(0.068)
CINC (Scaled)	-2.356^{***}	-2.242^{***}	0.918
	(0.444)	(0.458)	(0.834)
Constant	130.088^{***}	134.979^{***}	31.364^{***}
	(7.341)	(8.808)	(6.136)
Year Fixed Effects	No	Yes	Yes
Country Fixed Effects	No	No	Yes
Observations	4,159	4,159	4,160
R^2	0.162	0.223	0.526
Note:			*p<0.1; **p<0.05; ***p<0.01

Table B.2: Correlates of Collective Identity Projections with Reduced Antonym Set

	Collective Identity Percentile Rank		
	(1)	(2)	(3)
MID (Last 20 Years)	4.913***	4.668***	3.703***
	(1.034)	(1.006)	(1.085)
Total Trade Flows (Log)	−0.040	0.377***	0.502***
	(0.113)	(0.126)	(0.147)
Alliance	−1.778*	0.269	−0.629
	(0.917)	(0.966)	(1.431)
Common Language	2.238***	1.917**	
	(0.801)	(0.777)	
Distance (Log)	3.997***	5.599***	
	(0.797)	(0.799)	
POLITY Difference	−0.169***	0.020	−0.146*
	(0.056)	(0.056)	(0.076)
CINC (Scaled)	4.157***	2.930***	4.629***
	(0.454)	(0.466)	(0.933)
Constant	15.481**	−7.107	41.899***
	(7.508)	(8.963)	(6.864)
Year Fixed Effects	No	Yes	Yes
Country Fixed Effects	No	No	Yes
Observations	4,159	4,159	4,160
R^2	0.042	0.121	0.351

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

B.2 Antonym Pairs

This appendix presents the antonym pairs used to construct the friendship (Table B.3) and collective identity (Table B.4) dimensions.

Table B.3: Friendship Dimension Antonym Pairs

Friend Pole	**Enemy Pole**
friend	enemy
acquaintance	antagonist
ally	adversary
companion	combatant
partner	rival
teammate	opponent
confidante	nemesis
associate	competitor
collaborator	detractor

Table B.4: Collective Identity Dimension Antonym Pairs

Us Pole	**Them Pole**
I	he
I	she
me	him
me	her
we	they
us	them
mine	his
mine	hers
ours	theirs
my	his
my	her
our	their
myself	himself
myself	herself
ourselves	themselves

B.3 Country Preprocessing

Before estimating the models, I performed two preprocessing steps on the names of countries. First, I ensured that all mentions of countries with multiple names or those that underwent a name change were preprocessed to reflect a single name. Second, I concatenated the names of countries that had more than one word by replacing the spaces with an underscore. The full list of country preprocessing cases is presented in Table B.5. Any country not listed underwent standard preprocessing, which involved removing capitalization and punctuation.

Table B.5: Country Name Preprocessing

Name	**Preprocessed Name**
antigua and barbuda	antigua_and_barbuda
austriahungary	austria_hungary
bosnia and herzegovina	bosnia
burkina faso	burkina_faso
burma	myanmar
cape verde	cape_verde
central african republic	central_african_republic
ceylon	sri_lanka
costa rica	costa_rica
cote divoire	ivory_coast
czech republic	czech_republic
dahomey	benin
democratic peoples republic of korea	north_korea
dominican republic	dominican_republic
east germany	east_germany
east timor	east_timor
el salvador	el_salvador
england	united_kingdom
equatorial guinea	equatorial_guinea
federated states of micronesia	micronesia
german democratic republic	east_germany
german federal republic	west_germany
great britain	united_kingdom

Table B.5: Country Name Preprocessing (continued)

Name	Preprocessed Name
guinea bissau	guinea_bissau
guineabissau	guinea_bissau
holland	netherlands
ivory coast	ivory_coast
kyrgyz republic	kyrgyzstan
malagasy republic	madagascar
marshall islands	marshall_islands
new zealand	new_zealand
north korea	north_korea
north vietnam	vietnam
north yemen	north_yemen
papua new guinea	papua_new_guinea
peoples republic of china	china
republic of korea	south_korea
republic of vietnam	south_vietnam
rhodesia	zimbabwe
rumania	romania
saint kitts and nevis	saint_kitts_and_nevis
saint lucia	saint_lucia
saint vincent and the grenadines	saint_vincent_and_the_grenadines
san marino	san_marino
sao tome and principe	sao_tome_and_principe
saudi arabia	saudi_arabia
serbia and montenegro	serbia_and_montenegro
siam	thailand
sierra leone	sierra_leone
slovak republic	slovakia
solomon islands	solomon_islands
south africa	south_africa
south korea	south_korea
south sudan	south_sudan
south vietnam	south_vietnam
south yemen	south_yemen
soviet union	russia
sri lanka	sri_lanka
st kitts and nevis	saint_kitts_and_nevis
st lucia	saint_lucia
st vincent and the grenadines	saint_vincent_and_the_grenadines
surinam	suriname

Table B.5: Country Name Preprocessing (continued)

Name	Preprocessed Name
syrian arab republic	syria
the democratic republic of the congo	democratic_republic_of_the_congo
timorleste	east_timor
trinidad and tobago	trinidad_and_tobago
union of soviet socialist republics	russia
united arab emirates	united_arab_emirates
united kingdom	united_kingdom
united states	united_states
upper volta	burkina_faso
ussr	russia
viet nam	vietnam
west germany	west_germany
yemen arab republic	north_yemen
yemen peoples republic	south_yemen
zaire	democratic_republic_of_the_congo

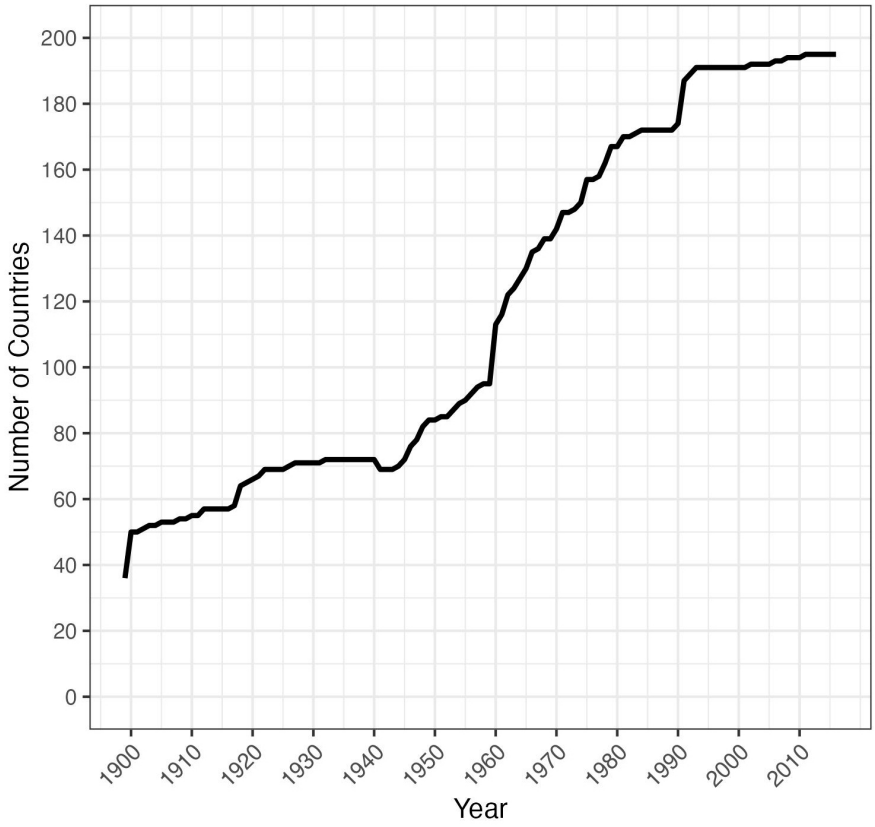
B.4 Summary Statistics of Words and Countries

Figure B.1 displays the total number of words in the sub-corpus for each session of Congress. Figure B.2 displays the total number of states in the world per year, according to the Correlates of War State System Membership (v2016) data set.[2] The increase in both corpus size and the number of states facilitates the estimation of projections for a greater number of countries in the later years in the data set.

[2]War Project 2017.

Figure B.1: Word Count Per Corpus

Figure B.2: State System Membership by Year



B.5 Placebo Tests

This appendix presents placebo tests for the friendship and collective identity dimensions using two sets of common words. Figure B.3 plots the annual friendship projection for the most common words from the 114th Congress, which ran from 2015-2017. Figure B.4 repeats the procedure for the collective identity dimension. Figures B.5 and B.6 replicate these plots, instead using the most common words from the 114th Congress that are not considered stop words.

Figure B.3: Historical Friendship Projection of Common Words

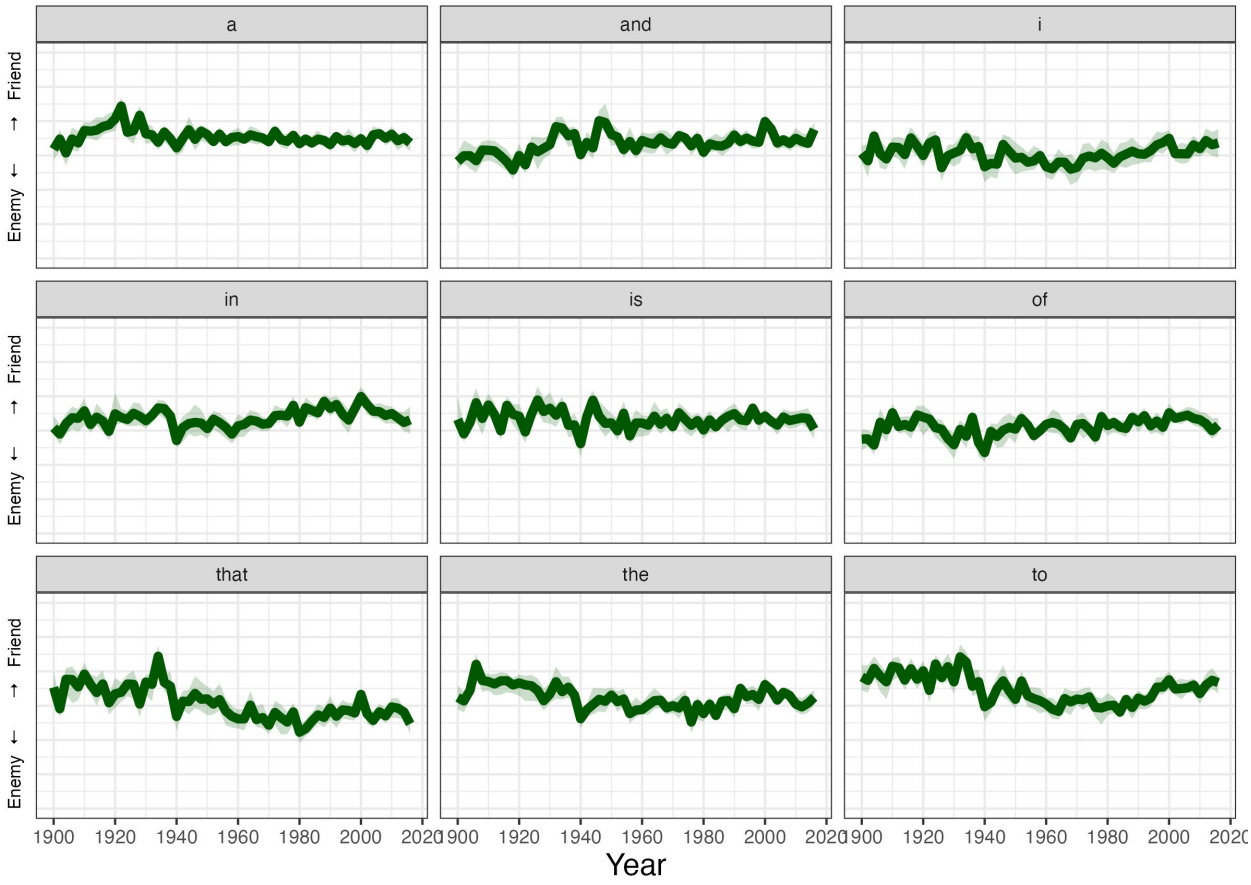


Figure B.4: Historical Collective Identity Projection of Common Words

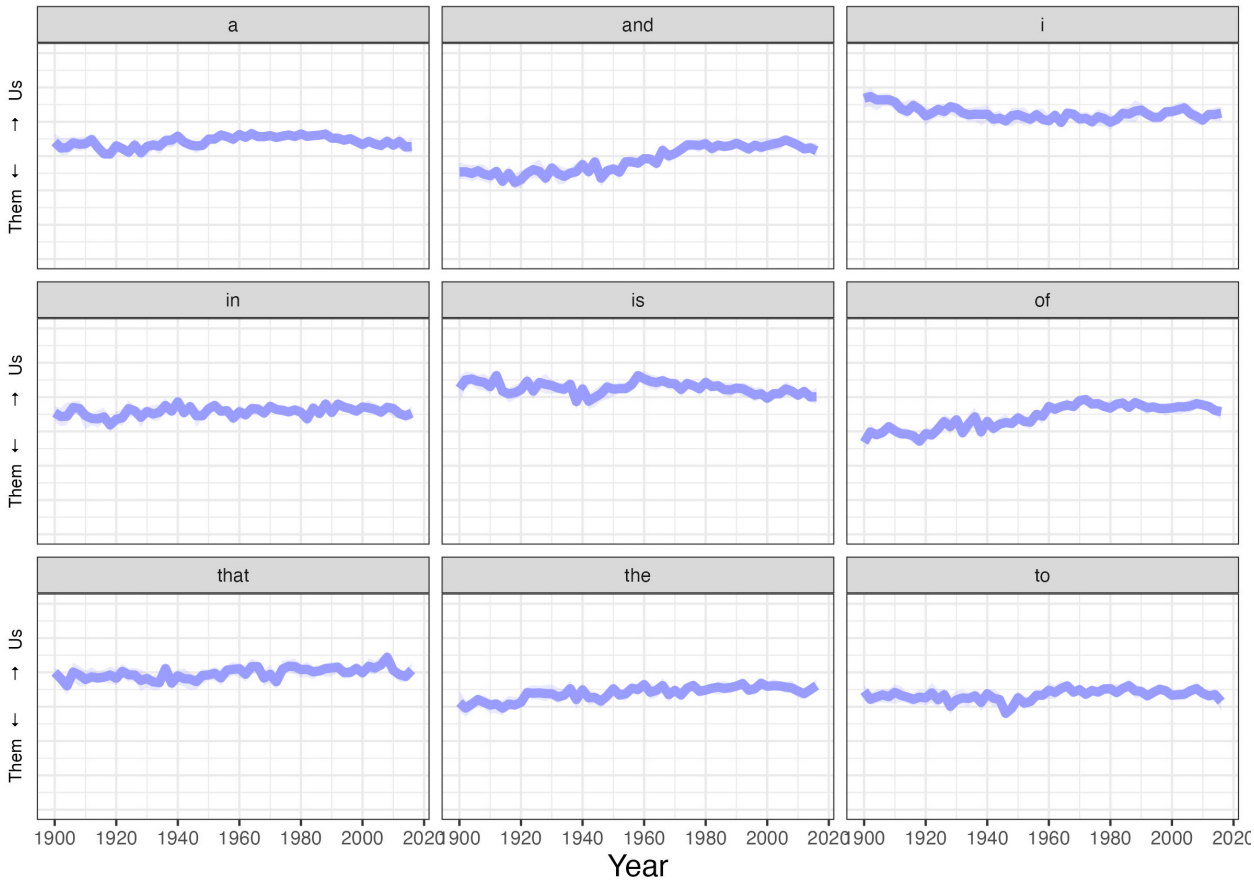


Figure B.5: Historical Friendship Projection of Common Non-Stop Words

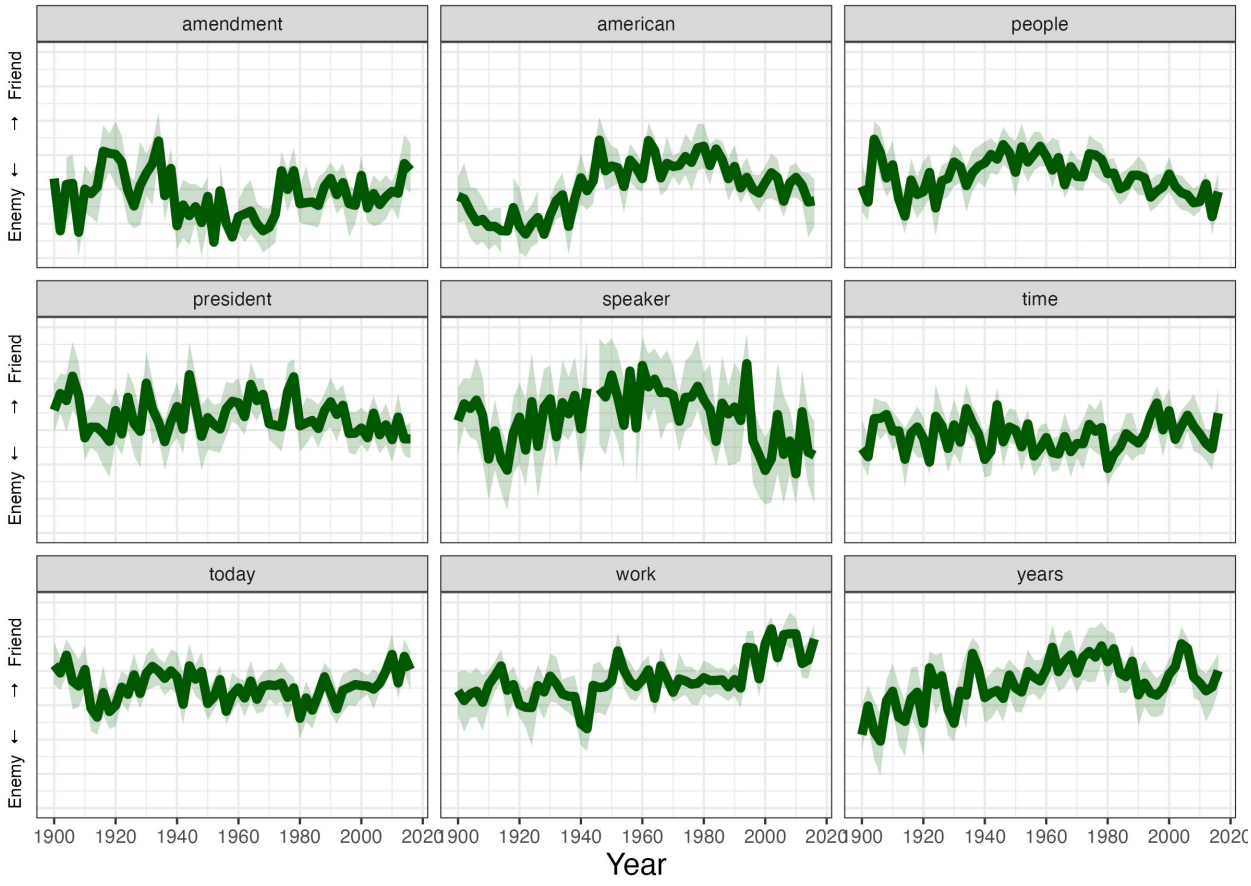
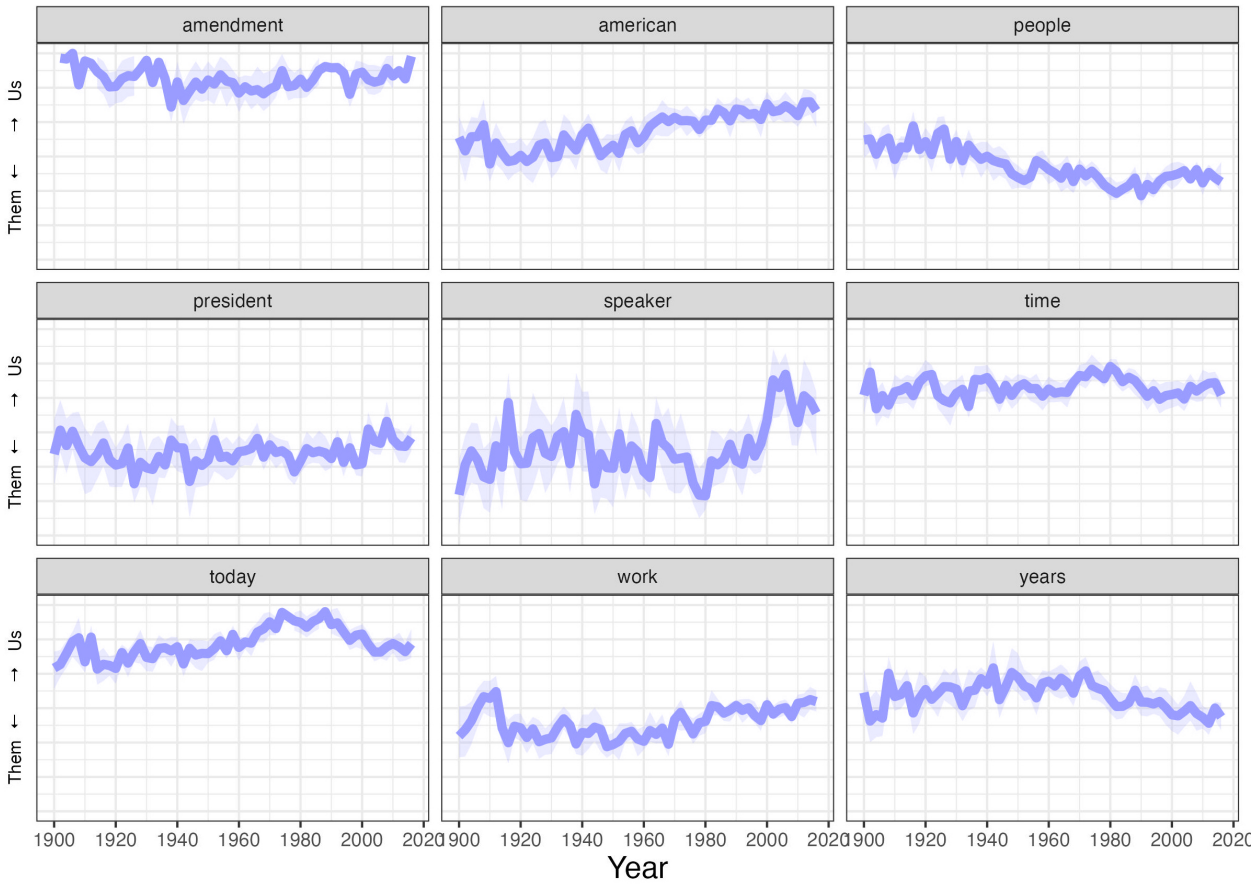


Figure B.6: Historical Collective Identity Projection of Common Non-Stop Words



B.6 Alignment Stop Words

To align the models, I use a set of stop words whose meanings I assume do not change. A stop word is only used to align two models if it exists in the vocabulary for both models. Table B.6 lists the stop words used for alignment.

Table B.6: Alignment Stop Words

a	become	downwards	happens	lest	none	regards	thats	value
able	becomes	during	hardly	let	nonetheless	relate	the	various
about	becoming	each	has	let's	noone	related	then	versus
above	been	eight	hasn't	like	nor	relatively	thence	very
according	before	eighty	have	liked	normally	respectively	there	via
accordingly	beforehand	either	haven't	likely	not	round	there'd	want
across	begin	else	having	likewise	nothing	said	there'll	wants
actually	behind	elsewhere	hello	little	notwithstanding	same	there're	was
after	being	end	help	look	novel	saved	there's	wasn't
afterward	below	ending	hence	looking	now	saw	there've	way
afterwards	beneath	enough	here	looks	nowhere	say	thereafter	welcome
again	beside	entirely	here's	lot	obviously	saying	thereby	well
against	besides	especially	hereafter	lots	of	says	therefore	went
ago	best	even	hereby	low	off	second	therein	were
ahead	better	ever	herein	lower	often	secondly	theres	weren't
aid	bettered	evermore	hereupon	ltd	oh	see	thereupon	what
ain't	bettering	every	hi	made	ok	seeing	these	what'll
albeit	between	everybody	his	main	okay	seem	thing	what's
all	beyond	everyone	hither	mainly	old	seemed	things	what've
allow	both	everything	hopefully	make	on	seeming	think	whatever
allows	brief	everywhere	how	makes	once	seems	third	whatsoever
almost	but	exactly	how's	many	one	seen	thirty	when
alone	by	example	howbeit	may	one's	send	this	when's
along	c'mon	except	however	maybe	ones	sent	thorough	whence
alongside	came	failing	hundred	mayn't	oneself	serious	thoroughly	whenever
already	can	fairly	if	mean	only	seriously	those	where
also	cannot	far	ignored	meantime	onto	seven	though	where's
although	cant	farther	immediate	meanwhile	opposite	several	three	whereafter
always	caption	few	immediately	merely	or	shall	through	whereas
am	cause	fewer	in	might	other	shan't	throughout	whereby
amid	causes	fewest	inasmuch	mightn't	others	should	thru	wherein
amidst	certain	fifth	indeed	million	otherwise	shouldn't	thus	whereupon
among	certainly	first	indicate	mine	ought	since	till	wherever
amongst	changes	five	indicated	minus	oughtn't	six	to	whether
an	circa	followed	indicates	miss	out	so	together	which
and	clearly	following	indicating	more	outside	some	too	whichever
another	come	follows	information	moreover	over	somebody	took	while
any	comes	followthrough	inner	most	overall	someday	toward	whilst
anybody	concerning	for	inside	mostly	own	somehow	towards	whither
anyhow	consequently	forever	insofar	much	owned	someone	tried	who
anyone	considering	former	instead	must	owning	something	tries	whoever
anything	contain	formerly	into	mustn't	owns	sometime	truly	whole
anyway	containing	forth	inward	my	particular	sometimes	try	whom
anyways	contains	forward	inwards	myself	particularly	somewhat	trying	whomever
anywhere	could	found	is	name	past	somewhere	twice	whose
apart	couldn't	four	isn't	namely	per	soon	two	why
appear	currently	from	it	near	perhaps	sorry	under	why's
appeared	definitely	further	it'd	nearly	placed	still	underneath	will
appearing	describe	furthering	it'll	necessary	please	stop	undoing	willing
appears	described	furthermore	it's	need	pleased	stopped	unfortunately	wish
are	describes	get	its	needn't	plenty	such	unless	with
aren't	describing	gets	itself	needs	plus	sure	unlike	within
around	despite	getting	just	neither	possible	take	unlikely	without
as	did	given	keep	never	probably	taken	until	won't
aside	didn't	gives	keeps	neverf	provide	taking	unto	wonder
at	different	go	kept	neverless	provided	tell	up	would
available	do	goes	know	nevertheless	provides	ten	upon	wouldn't
away	does	going	known	new	providing	tends	uponed	yes
awfully	doesn't	gone	knows	next	quite	than	upwards	yet
back	doing	got	last	nine	rather	thank	use	zero
backward	doings	gotten	lately	ninety	really	thanks	used	
backwards	don't	greetings	later	no	recent	that	useful	
be	done	had	latter	no-one	recently	that'll	uses	
became	down	hadn't	least	nobody	regarding	that's	using	
because	downward	half	less	non	regardless	that've	usually	