Platform Governance: The Transnational Politics of Online Content Regulation

Robert Gorwa

St. Antony's College

University of Oxford

Thesis submitted in partial fulfilment of the requirements for the degree of DPhil in International Relations in the Department of Politics and International Relations

73,500 Words

Trinity Term 2021

Abstract

Billions of people around the world use services like Facebook, Twitter, Instagram, and YouTube every day to access information, engage in conversation, and stay in touch with friends and family. These hugely profitable and popular platforms for user-generated content, operated by large multinational technology companies, have in the past decade created complex systems of private regulatory standards that govern online behaviour and have a significant impact on the social, cultural, and political lives of their customers around the world. Where these systems — which can be understood by International Relations (IR) scholars as an expression of private authority in global politics — were once tacitly accepted or ignored by state actors, governments have in recent years increasingly sought to shape the rules and practices deployed by platform companies through various strategies. In some cases, governments have sought to 'take back control' and re-assert state authority over this privately-managed domain, while in others they have opted rather to work directly with companies in more collaborative fashion. What explains the variation in how governments intervene in platform governance?

The thesis argues that how governments seek to shape, challenge, or contest private platform rule-making can be understood as either fitting in within a *collaborative* or a *contested* strategy. Building upon literatures from transnational regulatory politics, the thesis explains variation between these two strategies as the result of an interplay between three factors: domestic demand for change, the ability to supply that change (regulatory capacity and transnational or domestic institutional constraints on that capacity), and normative understandings of an actor's appropriate degree of policy intervention. The plausibility of the argument is demonstrated empirically through three qualitative case studies of key regulatory episodes (the German NetzDG, the Australian AVM Act, and New Zealand's Christchurch Call) in which different governments have deployed different strategies to affect platform rule-making.

Acknowledgements

Researching this topic and writing this DPhil made for the most fun and fulfilling — if occasionally challenging — four years of my life thus far.

My foremost appreciation goes out to my supervisors. Thanks to Lucas Kello for taking a chance on me and inviting me into the Department, for cultivating such a great network of digital politics students/outcasts, and for the big picture advice and guidance over the course of my research. A huge thanks as well to Thomas Hale for making the time to join the project at a critical juncture and helping me get across the finish line. I have learned a tremendous amount from you both.

In Oxford, I also benefited from the time, wisdom, and collaboration of Timothy Garton Ash and Rasmus Kleis Nielsen at various stages of my journey. Thanks to Vicki Nash, Abraham Newman, Duncan Snidal, Rasmus Kleis Nielsen, and Karma Nablusi for giving their time to examine this thesis from transfer and confirmation to the final viva. A big thanks as well to Taylor Owen and Heidi Tworek, who have supported me since my undergraduate days and have become indispensable colleagues and mentors. I would not have made it to Oxford without you.

There are so many wonderful people that I have met along the way that shaped this work through their community, conversation, and hospitality. In Oxford, the OII DPhil crew (especially Carl Öhman, David Watson, Nahema Marchal, and Corinne Cath), our Bullingdon Road family (Johanna Wetzel, Irene Fabricci, Lena Reim, Fuaad Coovadia, Ollie Ballinger, and of course Zoë Johnson), and the multi-cohort CTGA squad (Monica Kaminska, Valentin Weber, Jamie Collier, Max Smeets, Florian Egloff, James Shires) all deserve special mention for their friendship, camaraderie, and sage advice. Additional thanks to Adam Turowski and Joel Hart for the great music and food over the years, and to Michael Veale for the many (in most cases literally) instant messages and the intro to the UK tech policy pub scene.

In Berlin, I'm extremely grateful to Thorsten Theil and to Nicolas Friederici for hosting me at the Weizenbaum and Humboldt Internet institutes, respectively, and for the many wonderful conversations on platform economies, electronic music, and much more. Additional thanks to Anjo Peez for the introduction to Berlin, and to Torben Klausa, Niklas Rakowski, Amélie Heldt, Clara Iglesias Keller, Joao Magalhaes, and Christian Katzenbach for all of the conversation and collaboration around the topic of platform regulation in Germany and the EU. I've learned

loads from you all. Another special thanks goes out to the friends that have made Neukölln lockdown life manageable over these past two years, and have endured a slew of chats (and at times complaints) about the dissertation and dissertation life: Isabela Vera, Jérémie Bonnemort, Ben Overton, Austin Romeo, Chris Kardish, Daniel Jones, Jon Vrushi, Aly Oloo, Nils König. Additional thanks to my wonderful co-conspirators in all things platform governance, my friends-at-a-distance Robyn Caplan, Sonja Solumun, and Pranav Bidare.

This project could not have happened without the support of various funders, especially the Social Sciences and Humanities Research Council of Canada: I'm eternally grateful for the chance they took on a kid only two weeks into grad school, with a number of big ideas but little else under his belt. Additional thanks to Jill Cannon and the wonderful folks at the Canadian Centennial Scholarship Fund for generously granting me scholarships not once but twice, and to the Scatcherd European Bursary and the WZB Berlin for funding my research in Berlin. Finally, a huge thank you goes to all those who took the time to be interviewed for this project, helped me connect with interviewees, or assisted with the freedom of information request process. Special shout out to Nic Suzor, Wolfgang Schulz, Robert Madellin, Alex Pflaum, Julian Juarsch, Torben Klausa, Vera Weidmensch, and Chris Beall for their help in that regard.

More than anything, however, I am enormously indebted to my family. My parents have supported me in every possible way throughout my education, and have offered emotional and financial support while also trusting me to find my own strange path. I can't thank you both enough for believing in me, and for all that you went through so I could get here. *Dziękuję za wszystko. Bardzo Was kocham.*

A huge thanks go out as well to Conor and Jodie for the many delicious meals, the constant encouragement, and for always expressing so much interest in such an esoteric and niche topic. My deepest gratitude goes out to their incredible daughter, Zoë, for her constant love and companionship over these past four years. You have truly been a witness to this journey and all of its varying milestones, deadlines, and (sometimes) disappointments. Thank you for being not just the best partner/chef/wine taster/travel companion one could ask for, but also for keeping my rollercoaster on the tracks when things got bumpy. Thank you for making each day (even during a global pandemic) one to look forward to. I love you.

Acknowledgement of Integrated Publications

This thesis is entirely composed of original material that has not been published elsewhere, with a few particular exceptions.

The second half of of Chapter 1 (Introduction) features a literature review which integrates some content from single-authored articles I published in *Information,*

Communication and Society (Gorwa, 2019) and *Internet Policy Review* (Gorwa, 2019) following my DPhil transfer-of-status. The first half of Chapter 3 also contains a small amount of content from those two articles.

A earlier version of the case study presented in Chapter 4 (Germany) was recently published in *Telecommunications Policy* (Gorwa, 2021).

Chapter 5 (New Zealand), p. 167-168, features a brief description of automated content moderation systems from a co-authored article published in *Big Data and Society* (Gorwa, Binns, and Katzenbach, 2020, p. 8).

Contents

List of Figures

List of Tables

1 Introduction

Contents

1.1 Thesis Overview

On the 8th of January 2021, the sitting president of the United States had his Twitter account suspended. After years of using Twitter to spew vitriol and misinformation that energized his voting base, cannily influencing agendas and framing media coverage, President Donald Trump posted a tweet that Twitter's policy staff finally took as a bridge too far. Trump's apparent call to violence during rioting at the US Capitol in Washington led Twitter — quickly followed by Facebook, Google, Amazon Web Services, Shopify and multiple other technology companies — to

remove his widely-followed and influential accounts from their services.

For many commentators and global political figures reflecting on the decision, the episode appeared to signal a revolutionary moment for the exercise of private power in the 21st century, an extraordinary demonstration of the ability of new, powerful multinational information gatekeepers to influence public discourse, democratic transitions, and elections through their own private processes and policies (Jennen and Nussbaum, 2021; Roose, 2021). It was especially surprising when seen in the broader context of the rapid and transformative shift it represented: a few years earlier, Twitter, Facebook, and other major online platforms for user-generated content were still loudly and proudly proclaiming themselves 'neutral' platforms that merely facilitated public discourse (Halliday, 2012). Facebook CEO Mark Zuckerberg had repeatedly rejected the argument that his company should make what amounted to high-stakes political decisions over the acceptable bounds of speech or action online, infamously stating that Facebook should not become a global rule-maker and 'arbiter of truth' (McCarthy, 2020). Nevertheless, even before the Capitol Hill incident, platform companies had been gradually developing a complex, global socio-technical and quasi-legal architecture of 'content moderation' that influences what billions of people around the world could say, share, and discuss with their friends, crossing jurisdictional lines while impacting everything from national security to health policy (Gillespie, 2018; Moore and Tambini, 2018; Van Dijck, 2020).

The overlapping public and private regimes of platform governance — what I define in this thesis as the regulative institutions shaping how a platform can be used by its customers, and the political contestation that arises when other actors in turn seek to shape those rules and practices — have become a leading 21st century policy concern. Platform companies, according to some legal scholars, have become de-facto governors with tremendous implications for politics, society, and culture (Helberger, 2020; Kaye, 2019; Klonick, 2017; Suzor, 2019), perhaps making them some of the most powerful corporate entities in human history (Vaidhyanathan, 2018). This power has not gone uncontested: in May 2017, the world's first law to specifically set out rules for how platforms moderated the online environment navigated by

German residents was passed by the German Bundestag; it has since been followed by similar initiatives in countries like France, Singapore, Australia, Austria, Brazil, and numerous others, sparking what Flew and Gillett (2020) have called a new 'policy turn' in internet governance. In 2021, we appear to find ourselves at a critical juncture point: one where the rules for the contemporary internet are actively being re-written in a complex process of government, firm, and civil society contestation.

How should we understand these new processes of public-private regulatory contestation and ordering? For international relations (IR) scholars, these recent developments have major implications for enduring debates about globalization and the role of private actors within the international system. Although platform companies have yet to be closely examined by IR scholars, the rise of powerful private actors that conduct governance through their rule-setting and norm-making is not surprising when seen within the longer tradition of 'global governance' scholarship documenting 'governance without government' and the growing influence of non-governmental actors in many, if not most facets of global politics (Avant, Finnemore, and Sell, 2010; Dingwerth and Pattberg, 2006; Rosenau and Czempiel, 1992; Slaughter, 2004). In fact, a subset of this scholarship has for more than a decade focused on the increasing role that corporations, international non-governmental organizations, and other types of private actors play as 'private regulators' that can set transnational rules governing various aspects of international activity (Büthe and Mattli, 2011; Mattli and Woods, 2009). Private governance has been comprehensively studied by researchers looking at domains such as international finance, environmental protection, manufacturing, and intellectual property (Bartley, 2018; Falkner, 2017; Hall and Biersteker, 2002; Tusikov, 2017; Vogel, 2010).

The emerging political role of platform companies, however, appears to challenge some important existing assumptions within the global governance literature around how and why private actors are becoming more active, powerful, and authoritative transnational regulators. As Green (2014) has argued in a comprehensive historical analysis of private authority in world politics, private rule-making is expected to take on authority either (a) when states consciously and strategically delegate

regulatory functions to selected private actors, or (b) when other stakeholders in international politics willingly accept the rules made by enterprising private actors. The logic is that governments fundamentally structure politics, with the corporate form itself deeply contingent on rules and structures created by states (Ciepley, 2013); for this reason, international companies or other private actors, such as non-governmental organizations (NGOs), are subject to coercive power from governments seeking to get them to do their bidding (Drezner, 2008). Governments thus can — and frequently do — try and use "private actors as proxies in order to achieve desired regulatory outcomes" (Farrell, 2006, p. 353), a process that Green (2014, p. 7) calls "delegated private authority." Because corporations cannot deploy traditional methods of coercion against state actors to force them to adopt the rules that firms develop or desire,[1] companies functioning in liberalized free-market economies are free to develop their own regulatory standards or processes, and these forms of "entrepreneurial private authority" (Green, 2014, p. 6) — such as a voluntary supply-chain transparency certification system or eco-labels — can be willingly be adopted by other actors, granting *de facto* authority and influence to the actors that control that initiative.

The contemporary landscape of platform governance does not quite fit this bill. In contrast to the expectations of Green and others, the private power of today's platforms — which derives from their rule-making and rule-setting functions, their private regulatory orders shaping online expression, information consumption, and sociality (Kettemann, 2020) — was never formally delegated to today's most powerful platform multinationals by states. Instead, at least since 2015, a host of policy initiatives, mainly in the form of national legislation, as well as some regional co-regulatory efforts, have been developed by states in an apparent efforts to withdraw and curb private rule-making authority. These regulatory efforts, originating in countries like Germany, Australia, France, and Singapore, have

[1] Corporations of course can and do deploy various forms of 'voice' (such as lobbying, or threats to exit a jurisdiction and thus pass along economic costs) against governments, but these are persuasive, and not fundamentally coercive arguments in the same way as state power is. See Mikler (2018).

sought, either domestically or transnationally, to contest the legitimacy of platform content moderation and espouse the primacy of local domestic law. That said, not all countries are seeking to 'take back control' and place private platform rulemaking under public influence: in other high profile examples, New Zealand and the European Commission have sought to collaborate with companies, using soft law and voluntary, multi-stakeholder governance in an effort to achieve their policy goals in certain issue areas (Gorwa, 2019). In this important and emerging space of global technology policy, we still do not fully understand the conditions under which governments delegate to, contest, or accept platform authority.

1.1.1 Research Question and Argument

The goal of this thesis is to provide the first focused exploration of platform governance — of how private platform rule-makers interact with the governance stakeholders seeking to shape their rule-making — and to provide some insight into the specific form of private authority that platform companies appear to enact in global politics. To explore this empirically, I wish to ask the following research question:

> What explains the variation in how governments intervene in platform governance?

The core argument of this thesis is that governments demanding changes to the regulative institutions of platform governance pursue those changes through one of three governance strategies, which I develop by building on extant work on public-private authority in the transnational regulatory politics literature: *contested platform governance*, *collaborative platform governance*, and *neutral platform governance*. In *contested platform governance*, a governmental actor seeks to layer new, binding rules domestically or transnationally to override a private actor's policies and practices of rulemaking governance in a certain jurisdiction. In *collaborative platform governance*, governmental actors, rather than seeking to change their own domestic regulative structures and rules, work together with private actors to try and steer changes to the policies and practices of those private

actors, acknowledging private actor regulatory competencies and de facto granting them their private authority as rule-makers in a specific issue area. Finally, *neutral platform governance* occurs when a government either ignores, or tacitly accepts the existence of private platform authority over content moderation. In this strategy, governments either ignore platforms, or seek to exert pressure on them via the established channels of complaint managed and set up by platform companies (e.g., by seeking to influence firm policy representatives and spokespersons). Who writes the rules, and who enforces them? These three categories encompass a range of strategies through which governments seek — or not — to fundamentally shape the ways in which platforms create transnational private regulatory orders with social, political, and cultural ramifications.

Table 1.1: Typology of platform governance strategies.

	Neutral	Collaborative	Contested
Supplier of Rules	Firm	Multistakeholder	State
Type of Rules	Voluntary	Voluntary	Binding
State Rhetoric	Varies	Partnership, Collaboration	Control, Sovereignty

Building upon work looking at regulatory change in international political economy, especially perspectives grounded in historical institutionalism, I advance a simple 'power-plus' argument to explain the emergence of these three governance strategies in various contexts. Drawing upon power-driven accounts of global rule-making (Drezner, 2008) as well as more recent approaches that consider regulatory capacity more holistically (Newman, 2017), the argument is that the ability to engage in contested governance is a function of domestic and (in certain cases, transnational) power resources: if an actor seeks to contest platform authority, and has power to do so via binding rules, the actor will do so. Governments that do not have the power resources to layer new rules via domestic legal frameworks, and yet still seek to shape platform's content moderation practices, will need to resort to (less costly, but also less enforceable) voluntary, collaborative mechanisms. However, I argue, the decision to engage in collaborative versus contested governance is not purely explained by power — it depends on the character of the demand for

regulatory change, which, following historical institutionalist insights, is heavily shaped by the domestic and transnational institutional context (in particular, normative understandings of the acceptable degree of government intervention in the area of freedom of expression, and various institutional interdependencies that may constrain action). For this reason, under some conditions, state actors with the power resources required to contest private governance via binding rules will instead choose to pursue collaborative arrangements.

This argument, and the empirical material summoned to support it, have a number of implications for our current understanding of how governments interact with large, powerful technology multinationals. The central takeaway for an IR audience is that platform private authority displays some different characteristics than private authority in other privately governed domains of international politics, with important implications for debates about the growing relevance of private actors in the international system (Avant, Finnemore, and Sell, 2010; Drezner, 2008; Green, 2014). Platform companies provide an important case study into how new non-state actors are wielding political power in world politics, and having a significant political impact without formal consent or the formal adoption of their standards by other actors.

In particular, the thesis looks at the politics of *taking back control* from private actors wielding private authority, thus providing a preliminary case study into a potentially new and important form of private-public regulatory contestation in IR. For the broader audience interested in platform regulation and online content regulation more specifically, or the regulation of large technology companies more generally, the main takeaway is that regulatory politics are an important feature in helping shape the outcome of policy contestation between firms and governments. Platform regulation does not occur in a vacuum, or manifest as a simple reflection of shifting public opinion; it is shaped by varying degrees of government demand, constrained or enabled by a wealth of potential domestic and transnational institutions, and influenced by a normative environment that shapes both the scope of possible regulatory intervention as well as how far policymakers

are willing to go. These are all political factors missing from the existing and rapidly growing literature on the topic.

1.1.2 Research Design

To answer this research question, a careful empirical research design is needed. While the topic of this thesis is increasingly widely discussed in policy circles and the popular press, there still has been little focused empirical work that looks at the development of platform regulation in detail. There have been no larger-scale efforts to comprehensively map the universe of potential cases, in terms of formal and informal regulatory structures that have ramifications for firm behaviour in this space; nor have there been detailed case studies of the most consequential regulatory episodes that feature potentially illustrative examples of contestation or collaboration between firms, governments, and other governance stakeholders.

This thesis takes a mixed-methods approach. I begin with a large scale longitudinal overview of the regulatory universe, synthesizing an original dataset of formal governance instruments as well as of private governance institutions in an effort to provide the most comprehensive overview of platform content regulation that has been assembled to date. This data collection used the leading repositories of formal regulation in intermediary liability policy (the Stanford World Intermediary Liability Map), as well as novel data on private regulatory organizations for online content collected and coded following best practices outlined in the largest existing dataset of private regulatory organizations, the 'Transnational Public-Private Governance Initiatives in World Politics' dataset (Westerwinter, Abbott, and Biersteker, 2021). By combining these two approaches, the thesis thus seeks to provide the first comprehensive macro-level overview of 'platform regulation' as a historical and political object of study.

After having described and delineated the universe of possible regulatory initiatives to be assessed, the central empirical contribution of this thesis involves the analysis of a range of hereto under-explored cases, key regulatory episodes leading to a new regulatory arrangement. Case study research is commonly used in the social

sciences to fulfill a range of functions, from providing deep, single-case analysis of a certain phenomenon to broader, multi-case comparison and generalizability across a universe of cases (Gerring, 2006). Case studies can feature either large-N quantitative or small-N qualitative designs, each of which serve different aims and have different strengths and limitations (Seawright and Gerring, 2008). As I deal with a relatively small number of regulatory initiatives for harmful content, and because each case is complex, context dependent, and depends on causal mechanisms that are not easily operationalized quantitatively, I opt for a small number of focused comparisons of individual cases to support my argument.

These case studies draw upon a variety of data, the two most important of which are qualitative interviews with governance stakeholders and new primary documents obtained via freedom of information requests. Despite their limitations, qualitative interviews with key policymakers and other elites have long been a common method used by political scientists (Leech, 2002). The aim of interviews, which are based on a conversation where the researcher asks questions and the respondent answers, can be to gather facts and information that may not be in the public domain, or which public documents might misrepresent (Tansey, 2007). For this thesis, I conducted fifty-two interviews with policymakers and regulators, former and current company employees, members of civil society, and others involved in the key regulatory episodes that are the focus of the case study chapters. My broad set of participants was selected using theoretical sampling, where after outlining the criteria for the population that I was interested in examining, I made initial inroads into a population and used access and contacts to obtain future interviews based on a snowball process (Van Evera, 1997). More information about the interview process, including the project's research ethics approval, participants, and the attribution of source material can be found in Methods Appendix A.

Secondly, alongside the close reading of publicly available policy documents, press releases, public testimony, and public interview transcripts that are the bedrock of academic desk research, I draw upon a new trove of publicly released primary documents obtained through the freedom of information requests made by myself and

other researchers, activists, and journalists. The thesis uses more than a thousand pages of new documents obtained from ten freedom of information access (FOI) requests levied to various government bodies, including the European Commission, the German Ministry of Justice and Consumer Protection, the Australian Attorney General's office, and the New Zealand Ministry of Foreign Affairs and Trade. I used these requests in an effort to provide an inside look into decision-making processes, as well as the internal deliberation around certain policy initiatives, that can otherwise be difficult to obtain. I sought to use these deliberative primary documents in an attempt to remedy some of the commonly accepted shortcomings of interviews, including the tendency of policy episode participants to portray themselves in a positive light, and to contextually situate and triangulate my findings (Silverman, 2015). More information about the FOIA process, and the way I formulated, publicly archived, and analyzed these primary documents is available in Methods Appendix A.

1.1.3 Contributions

My argument has implications for how the regulation of technology companies is currently understood by scholars, policymakers, and the public, and I hope that it can make a contribution to the global governance and regulatory politics literatures addressing emerging technology policy domains as well as the more specific 'platform studies' conversation that my work has been situated within.

Firstly, my work makes a theoretical contribution to IR by applying and extending insights from transnational regulatory politics scholarship, drawing upon work on private authority, regulatory and institutional change, and extending it into a hereto unexplored domain. The thesis looks at the politics of *taking back control* from private actors wielding private authority, providing a novel theoretical twist on the existing literature. Looking at this type of private authority in world politics demonstrates both the importance of power and powerful actors working within largely established models of international regulatory politics, and also emphasizes the importance of norms around free expression and constraints on understandings

of appropriate policy action. It additionally highlights for IR scholars the ongoing importance and political salience of certain private standards, but also demonstrates that powerful private actors can (and do) in many ways resist state coercion when their rule-making is contested. While the thesis focuses on the necessary first condition for understanding contemporary platform regulation — the conditions under which government actors intervene and engage in contestation — and not the corollary question of under which conditions are those interventions truly successful in fomenting long-term change that meets actor preferences and goals, a major takeaway of this project is that in some (highly technical, globalized, and high stakes) political domains, state power and interest alone appears to be insufficient for achieving major reforms to the privately-developed and managed status quo.

I believe that my thesis may also provide a contribution to other domains where the emergence of private authority does not clearly track the pattern established in more institutionalized and formalized governance arenas. I tell a story of a policy arena where private authority grew unexpectedly and quietly, in a manner tacitly tolerated or ignored as a low-salience policy issue without formal delegation or agreement from state actors, setting the stage for potential clashes with state interests once the impact and salience of private platform rule-making grew. Eventually, the private regimes created by platform companies came to be framed in various policy discourses as a potential threat to state interests, with governments working since 2015 to influence these private regulatory standards with a variety of tactics, which have yet to be explored in-depth by IR scholars; the thesis contributes a case study of this emerging and important policy area to transnational regulation scholarship, and helps test the boundaries of some of the existing theorizing in this area by applying it to this fast-moving, technologically complex domain.

The questions at hand are highly topical, having been discussed heavily not only in public discourse, but also in a large and growing body of interdisciplinary academic scholarship outside of political science. I intend this thesis to also contribute to the body of insights into power of platform companies in the literature that has been produced by digital media and communication scholars (Hargittai, 2007; Langlois,

2013; Nieborg and Poell, 2018; Weltevrede and Borra, 2016), as well as legal scholars who have long been on the forefront of examining private ordering and power online (Bietti, 2021; Klonick, 2017; Lessig, 2009). While existing accounts looking at the role of platform companies in contemporary politics tend to come from a sociological, philosophical, or cultural studies tradition (Bucher and Helmond, 2018; Vaidhyanathan, 2018; Van Dijck, Poell, and Waal, 2018), emphasizing the cultural and social aspects of digitized globalization, the conceptual approach in this thesis foregrounds an explicitly political and policy oriented approach that I argue is missing from existing 'platform studies' scholarship (Bogost and Montfort, 2009; Gillespie, 2010; Gillespie, 2014; see also my contribution to Gillespie et al., 2020). It seeks to provide an awareness of how government interests interact with firm motivations, and to conceptualize the institutions of platform governance as constantly shaped by a process of political contestation across different actor groups.

Perhaps the most significant contributions of the thesis, however, are its empirical contributions that aim to improve our understanding of the workings of platform governance in practice. This thesis combines comparative macro-level insights gained from my mapping of the regulatory landscape with the analytical lenses afforded by a transnational regulatory politics approach, providing new descriptive insights into the state of private regulatory organizations and informal regulatory arrangements in platform governance. The case studies examined in this work have, to date, either been largely unexamined in peer-reviewed outlets or have not yet been analyzed outside of a purely legal and textual framework. These cases thus provide new analyses of important regulatory episodes, and thus may be of interest to scholars, journalists, and policymakers seeking to better understand the process leading up to a certain outcome. I hope that the case studies provided here can help provide the first step towards a comprehensive comparative policy analysis framework that can better explain regulatory outcomes in different contexts across various technology policy issue areas.

Finally, the thesis makes a small empirical contribution to future research undertaken in this area by publicly archiving its key empirical material collected

via freedom of information requests, so that it is searchable and accessible for future scholarship. I hope to expand upon this methodological contribution to IR and regulatory scholarship by continuing to refine and elaborate a strategy for undertaking research with freedom of information requests in a transparently cite-able, robustly archived, and reproducible manner.

1.1.4 Chapter Structure

This thesis is structured in seven chapters. First, the rest of this introductory chapter presents a few key definitions deployed in this thesis and offers a more comprehensive overview of the relevant literature that this thesis contributes to, discussing emerging work on governance in digital media, political communication, and 'internet studies' circles, as well as the larger body of relevant scholarship in global governance that deals with corporate actors.

Chapter 2, the theory chapter, presents the conceptual model used to answer the central research question and to structure the ensuing analysis of change in platform governance. After an overview of various understandings of private authority in studies of global regulation, the chapter outlines a tripartite typology representing state strategies for pursuing their preferences in platform governance, and provides an argument seeking to explain under which conditions change is to be expected.

Chapter 3 is a data-driven chapter that offers an overview of the general evolution of platform regulation since its inception with the rise of user-generated content platforms in the mid-2000s. It seeks to provide overview of the global universe of formal and informal regulatory mechanisms that affect how platforms govern harmful content, drawing upon data from the Stanford World Intermediary Liability Map as well as newly compiled data on private regulatory organizations active in the platform regulation space. The goal of the chapter is to provide a universe of potential cases for the rest of the thesis, and it concludes with a discussion of case selection for the case study chapters.

The case studies begin in Chapter 4, which focuses on a regulatory episode involving the fight to impose standards for how platform companies conducted their

online content moderation in Germany from 2015 to 2018. The chapter discusses the German Network Enforcement Law (NetzDG), and outlines how German policy entrepreneurs, motivated by domestic and electoral factors, sought to layer new domestic rules on top of the existing European platform liability regime. The chapter demonstrates how they were able to supply the demanded rules despite not only significant opposition from industry and global civil society, but also the institutional constraints of a European Union set up to maintain regulatory harmonization.

Chapters 5 and 6 discuss the diverging regulatory responses of Australia and New Zealand in the aftermath of the March 2019 Christchurch attack. These cases provide an example of two neighbouring countries with close social, economic, and political ties seeking to respond to the same event — and achieve the same stated policy goal of regulating terrorist content on major social media platforms — with two very different strategies, in effect providing a policy experiment that helps illustrate the factors that motivate the realization of contested versus collaborative private governance.

In Chapter 5, I discuss why the New Zealand government opted for a collaborative governance approach in partnership with international platform companies, orchestrating an international voluntary regulatory initiative called the Christchurch Call. Despite having the power resources sufficient for passing domestic legislation that would have produced a contested platform governance strategy, I argue that New Zealand was motivated by normative considerations (in particular, a less interventionist, more *laissez-faire* attitude around the appropriate role of government in curbing free expression online) to pursue a collaborative governance strategy instead.

In Chapter 6, I argue that the Australian government, in contrast to its neighbours in New Zealand, was driven by domestic preferences right before the end of a parliamentary session to demand new rules. The Australian executive used their domestic power resources to initiate a contested, security focused regulatory approach that substantially increased liability for platform companies operating

in the region, with the Australian executive unconstrained by normative concerns or other major institutional barriers.

Chapter 7, the conclusion, summarizes the findings of the thesis overall, outlining some of the potential policy ramifications of its arguments, directions for future research, and a more detailed discussion of its contributions and limitations.

1.2 Key Concepts and Literature

The topic of this thesis spans many disciplines and includes much jargon. The remainder of this chapter seeks to lay a basic foundation by providing an overview of the main concepts and literatures upon which this thesis builds. First, I begin by discussing what platforms are, their characteristics, and past efforts to classify them. Second, I discuss the growing literature on 'platform governance' in digital media and communications studies, situating that conversation within broader literatures looking at multinational and transnational corporations in international political economy. Finally, I conclude with a discussion of why platforms matter for international relations scholars and political scientists, and an identification of the major gaps in both the more general global governance/transnational regulation literature that does not substantively engage with platform companies as well as the more focused interdisciplinary internet and media governance literature that does.

1.2.1 What is a Platform?

"Platform" is a particularly ambiguous term that is used differently by various scholarly communities (Andersson Schwarz, 2017). In particular, there are important variations in the ways that platforms are understood in computer science, economics, digital media studies, and legal scholarship.

The earliest usage of the term in its computational sense likely began in the 1990s in California, as software developers began to conceptualize their offerings as more than just narrow programs, but rather as flexible "platforms" that enabled code to be developed and deployed. A 1995 pamphlet by Sun Microsystems described operating systems like Linux, Mac OS, or Microsoft Windows as platforms,

generally understood as infrastructures upon which code can be deployed (Bogost and Montfort, 2009). The term did not take off, however, until the early 2000s, when a new crop of technology entrepreneurs found the old notion of a flexible computational "platform" particularly compelling in the so-called "Web 2.0" era of user-generated content. Mark Andreessen, a technology entrepreneur who created the Mosaic and Netscape web browsers, outlined platforms as follows:

> Definitionally, a "platform" is a system that can be reprogrammed and therefore customized by outside developers — users — and in that way, adapted to countless needs and niches that the platform's original developers could not have possibly contemplated, much less had time to accommodate (Andreessen 2007, Quoted in Bogost and Montfort, 2009, p. 4).

Platforms therefore came to represent what the legal scholar Jonathan Zittrain has called "generativity" — the ability for an object to be built on and adapted in ways beyond its initial purpose, and perhaps even beyond what its creators could have imagined themselves (Zittrain, 2008, p. 2). According to such definitions, virtually any hardware or software that embodied this ideal could be considered a platform, with some going as far as to refer to the internet itself as the ultimate platform (O'Reilly, 2007). However, the term became more more closely associated with online software, and specifically with a core feature called the Application Programming Interface, or API: a tool that allows external software developers to generatively draw upon data, execute commands, and build upon the online service that provided the API (Helmond, 2015). Writing in 2007, Andreessen states that the Firefox web browser, the image sharing website Flickr, the payment service Paypal, and the recently founded social network Facebook are all example of platforms, presumably because they all have APIs. "If you can program it, then it's a platform. If you can't, then it's not," he argued (Andreessen, 2007, n.p.).

As more and more of these generative online services were founded in the early 2000s (and more and more of them became viable businesses, even after the dot-com crash), it was only a matter of time until the economists started paying attention. Focusing less on APIs and the other infrastructural elements which characterized

platforms from a computational standpoint, the pioneering work of economists like Jean Tirole honed in on the ways that various platforms profited by bringing multiple parties together. Platforms were defined here as technologies or services that mediated interactions and relations between two or more parties, with their core feature being their effective identity as multi-sided markets (Rochet and Tirole, 2003). For instance, Rochet and Tirole noted that the makers of operating systems such as Windows had to bring together both the software developers who would create applications for it and the regular users who would use those applications (and therefore the operating system), while the providers of "credit card platforms" had to get both businesses and users to buy in (businesses, to purchase and use the credit card terminals; ordinary consumers, to use the specific type of credit card) (Rochet and Tirole, 2003, p.1014). One of the key insights of this literature on platform competition is that the platform operators can generally choose to make their profits on only one of these two market sides (for Microsoft, they chose to sell their operating system to users, and to offer costly perks to developers to incentivize them to build applications, while a company like Visa chose the opposite approach, offering their service at a subsidized price point to the public in an effort to make their money off of the businesses paying fees for each transaction).

Since their emergence in the late 2000s, a number of American companies have captured the majority of the global market for services commonly called 'social networks' (boyd and Ellison, 2007). These firms, most notably comprising of Facebook (2.7 billion global users), Youtube (2.2 billion), Instagram (1.2 billion), Twitter (353 million), as well as more recent entrants like Snapchat (498 million) and the new Chinese company TikTok (689 million; all estimates from Statista, 2021), provide internet-protocol based applications that allow customers — usually able to join via a free subscription — a way to upload and share their content (text, video, images, audio) to the service. They can then navigate and access the content shared by others (either acquaintances or strangers), and interact in a structured quasi-social manner with that content (Burgess, Marwick, and Poell, 2018). As most of these services work off of a multi-sided business model, where

the companies allow advertisers to target customers on the 'social' side of the network with proprietary data collection and targeting infrastructures (Gawer and Cusumano, 2014), provide a surface upon which third-party developers can build functionality (Bogost and Montfort, 2009), and ostensibly champion the ideological assumptions of 'openness and neutrality' (Gillespie, 2010), they have come to be commonly called 'platforms' in public discourse (Jørgensen, 2019; Srnicek, 2016; Van Dijck, Poell, and Waal, 2018).

The economic notion of the platform as a space which brings together multiple parties proved to be immensely influential, and injected the term with symbolic, strategic, and political weight. Part of the shift came as young technology companies like Google, Ebay, Yahoo, Facebook, and YouTube grew rapidly and started running into thorny legal issues. In one early incident, Yahoo was sued by a French citizen who objected to the online sale of Nazi memorabilia via Yahoo's auction service, an activity illegal in France (J. L. Goldsmith and Wu, 2006). In another, Google came under pressure to remove links to certain content, such as anti-semitic webpages that occupied the top spots for searches for "jew" or "jewish" (Grimmelmann, 2008). Through a multitude of similar scenarios, key questions about the legal responsibility of these technology companies emerged: should search engines and other online products be held liable for the content that they presented to users? Or were they merely neutral "intermediaries" that brought multiple parties together (Ardia, 2009)? As these key legal questions were being settled, it became increasingly useful for certain technology companies to brand themselves as "platforms" that facilitated access to user-generated content, but did not create it, and therefore should not be held liable for it. This was more than an empty rhetorical move: as Gillespie and others have argued, the language used to describe technology has always been tremendously important:

> As society looks to regulate an emerging form of information distribution, be it the telegraph or radio or the internet, it is in many ways making decisions about what that technology is, what it is for, what sociotechnical arrangements are best suited to help it achieve that and what it must not be allowed to become. This is a semantic debate as much as anything else: what we call such things, what precedents we

> see as most analogous and how we characterize its technical workings drive how we set conditions for it (Gillespie, 2010, p. 355-356).

YouTube press materials began in 2006 to refer to the service as a "platform for people to share their videos around the world" (Gillespie, 2010, p. 352). Facebook CEO Mark Zuckerberg initially referred to his product as a "utility" — reflecting his goal to provide an understated service that was a generally useful and essential part of people's everyday lives — before being forced to re-brand by company lawyers, who warned that utilities were tightly regulated in many countries (Fisher, 2018; Hoffmann, Proferes, and Zimmer, 2018). The contemporary usage of the word platform therefore cannot be separated from the legal landscape that most major technology companies have had to navigate — it increasingly became a term used to occupy a strategic niche in American telecommunications law, walking the line on important pieces of intermediary liability regulation, such as Section 230 of the Communications Decency Act (Gillespie, 2018). By branding their services as platforms, technology companies were able to portray themselves as open, neutral conduits which did little more than provide a good that could be used in countless ways by various parties (Gillespie, 2010; Van Dijck, 2013).

With this evolution, the usage of platform expanded far beyond its original, computational definition. Today, services like Airbnb and Uber are not just commonly referred to as platforms, but themselves have adopted the term as a shield against lawsuits and regulation. Uber, for instance, is effectively a taxi company but positions itself as a mere software platform that matches drivers with app users, avoiding traditional municipal restrictions on taxis (Calo and Rosenblat, 2017; Rosenblat, 2018). Similarly, Airbnb provides accommodation like a hotel chain, but brands itself as an impartial platform matching users and homeowners, avoiding existing restrictions on hospitality provision in cities (Edelman and Geradin, 2015). The language not only is a clever public relations move, hinting at user empowerment and the apolitical provision of services, but also pragmatically helps deflect political and legal responsibility. In effect, as Gillespie has argued, contemporary "Platforms are 'platforms' not necessarily because they allow code to be written or run, but

because they afford an opportunity to communicate, interact or sell" (Gillespie, 2010, p. 351). With this in mind, some of the latest scholarship defines platforms not as a form of programmable technical infrastructure, but rather as a specific form of data-driven business model (Bietti, 2021; Cohen, 2019; Srnicek, 2016).

As the definition of platform has become more business oriented, the word has taken on a dual meaning, referring not only to services being provided (i.e., Facebook.com is a platform) but also to the provider of the service themselves (Facebook Inc. is commonly also called a platform). Exacerbating the confusion even further, it is generally accepted that a "platform" like Google LLC can be composed of many different platforms (e.g., Google Search, YouTube, Google Cloud Services) (Srnicek, 2016). There are multiple competing definitions, which lead to all sorts of different analyses, and an important distinction must be made between *platforms*, which I define for the purposes of this thesis as *the data-driven, internet-enabled products that mediate relationships between two or more parties*, and *platform companies* (quite simply, the technology companies that own and operate platforms).[2]

Defining a Platform Universe

Part of the challenge with the platform term is that different services fit under that broad umbrella. For instance, a report prepared for the European Commission in 2017 suggested that there were more than two hundred relevant platform companies operating in Europe (Fabo et al., 2017). By 2020, an explanatory memorandum published by the Commission in a proposal for new platform-related regulation was arguing that there were more than ten thousand (European Commission, 2020). These can vary widely in terms of size and core business model, ranging from industrial platform operators (e.g. Siemens) to 'gig economy' or 'sharing economy' platforms like AirBnb, Uber, Taskrabbit, and others. All platform companies must set some rules that determine who can use their platform services and how those

[2]This definition has three notable aspects: a) it acknowledges technical features while noting that contemporary platforms are at their core products designed to generate maximum profit for the companies that operate them; b) it acknowledges that platforms are never neutral, and that "a platform is a mediator rather than an intermediary" (Van Dijck, 2013, p. 29); and c) it acknowledges that platforms are multi-sided markets that structure relationships between a number of different customers.

platforms are used, but these evidently vary in their complexity, political salience and public significance. If possible examples of platform companies include not just Facebook, Google, Uber, AirBnb, Amazon, Twitter, Microsoft, Apple, and others but also prominent non-American platforms like WeChat, Baidu, JD, Alibaba, and TikTok, we need a mechanism to distinguish between platforms and classify them into categories. What are their analytical properties, and importantly for political scientists, how do they vary in their global political impact and importance?

Given that there is no widely-accepted definition of precisely what a platform is, it is unsurprising that there is no agreement on exactly how platforms should be classified. Management researchers have been writing about the platform business model for more than a decade, and were the probably the first to propose explanatory typologies of different types of platforms (D. S. Evans, Hagiu, and Schmalensee, 2008; Gawer, 2011). One of the more ambitious recent efforts by Evans and Gawer has built upon this literature in an effort to identify, classify, and analyze every one of the world's significant platform companies (as defined by a market valuation of more than one billion US dollars). The authors define three categories relevant here: transaction platforms, which are multi-sided markets and intermediaries that bring multiple parties together; innovation platforms, which provide the building blocks for future entrepreneurship (indirectly echoing the original, computational idea of a platform); and integrated platforms, which combine features of the two (P. C. Evans and Gawer, 2016).

Table 1.2: Management-oriented typology of platforms, Evans and Gawer (2016).

Type	Examples	Key Feature
Transaction Platforms	Uber, AirBnb, Tencent, Spotify	bring together multiple parties and facilitate transactions between them
Innovation Platforms	Microsoft, Salesforce, Intel	provide a foundation on top of which other firms can build
Integrated Platforms	Google, Facebook, Alibaba	combining features of both transaction and innovation platforms

The authors identify 176 total platforms that meet their broad criteria, with 82 in Asia and 64 in North America. But the difficulties inherent in their scheme become immediately apparent: services like Google Search or Amazon Web Services and Amazon Marketplace are all counted as individual instances of transaction platforms, while their parent businesses, Google and Amazon, are double-counted as integrated platforms. This approach is too general to be analytically useful: if a company has some form of multi-sided market but also runs an API ("innovation") it becomes an integrated platform. Evans and Gawer state that all of the most valuable technology companies (Facebook, Google, Amazon, Apple) are transformative because they are integrated platforms, but do not elaborate on the broader implications that might follow from this argument.

Other efforts have attempted to categorize firms by their core purpose or industry. "Social media platform" has become a popular term of art, used by journalists as well digital media and communication scholars (Van Dijck and Poell, 2013, p.3). Gillespie's recent definition of the social media platform provides many examples, including not only "Facebook, YouTube, Twitter, Tumblr, Pinterest, Google+, Instagram, and Snapchat... but also Google Search and Bing, Apple App Store and Google Play, Medium and Blogger, Foursquare and Nextdoor, Tinder and Grindr, Etsy and Kickstarter, Whisper and Yik Yak" (Gillespie, 2018, p. 255). The first eight listed seem to be clearly social networks along established definitions (boyd and Ellison, 2007), but do app stores really have the same affordances as social media? What about Etsy and Kickstarter, a website for selling handmade arts and crafts, and a website to fundraise money for projects or causes, respectively? Should a service like Twitter be placed alongside Facebook and YouTube, despite having a significantly different user-base and affordances (Bucher and Helmond, 2018)? This type of list-based approach is limited not only by definitional challenges, but also by the fact that the larger platform companies are generally composed of many different platforms and it can be difficult to unpack out exactly what their core service or industry is. Alphabet, the holding company set up in 2015 to separate Google from some of its acquisitions, operates not only the world's most influential

search engine (Google Search) and video-based social network (Youtube); it also offers enterprise cloud hosting (Google Cloud), email services for individuals and businesses (GMail), and a major mobile operating system (Android), along with phones (Pixel), smart speakers (Google Home), and other home devices (such as the Nest thermostat). Therefore, it is not clear if Google is a social media platform — should there perhaps be a separate category of "search platforms"? Many of the largest, and most influential technology companies seem to resist classification (after all, their differences from the competition helped make them dominant in the first place), now posing a key challenge for regulators, who struggle to figure out exactly what the "platform industry" is and what industry-wide policy should look like. In that sense, some might argue that each one of the largest platform companies (what some refer to as GAFA: Google, Apple, Facebook, Amazon) represents a unique platform type (Barwise and Watkins, 2018).

In *Platform Capitalism*, Srnicek suggests a different approach, providing a typology that classifies platform companies based on what he perceives to be their core business model (Srnicek, 2016). He outlines three broad categories relevant for this thesis: advertising platforms, cloud platforms, and lean platforms. While Google and Facebook may initially seem to be in different markets (search vs. social networking), Srnicek argues that they are unified by their overarching goal to collect user data and monetize it by selling targeted advertising. Although both companies profile their users in different ways, it is the core feature of both of their business models: in 2016, Google reported that 89 per cent of their revenues came from advertising; for Facebook, that number was 96.6 per cent (Srnicek, 2016). Srnicek suggests that companies reliant on advertising are likely to share a number of features that bring them together, such as their operation of social networks or other forms of online services that are especially data rich and provide opportunities to collect valuable (and sensitive) personal information.

His second category, cloud platform, is fundamentally about providing computing infrastructure to businesses and individuals. According to Srnicek, Amazon is the most significant cloud platform, an argument that appears ludicrous on its surface

— after all, everyone knows that Amazon is a retailer and marketplace, world-renowned for its sprawling, high-tech physical infrastructure for warehousing and logistics. But Amazon Web Services (AWS), which provides hosting, servers, and other online infrastructure on demand, has rapidly grown to be a huge part of the company's business. In 2017, AWS produced the majority of Amazon's profit for the first time, lending credence to the argument that this has become Amazon's core business model: by operating on 25-30 per cent profit margins, it now subsidizes the entire retail operation, which operates on razor slim 2-3 per cent margins in North America and at a loss in Europe (Srnicek, 2016). Cloud platforms can have generative aspects, offering the computing and hosting capacity which much of today's platform economy relies on, and when AWS or Microsoft Azure goes down, so do a significant proportion of some of the internet's most popular websites. As firms like Uber have been able to flexibly scale to new markets by relying on the just-in-time computing capacity offered by cloud platforms, Srnicek argues that cloud platforms form the backbone of the contemporary platform economy.

Finally, Srnicek describes lean platforms: companies which favour growth before profit and hyper-outsource all they can, including not only workers, training, and maintenance costs, but also their entire cashflow, which is effectively outsourced to venture capital firms (Srnicek, 2016). Being lean is not just a feature, but it is the central tenet that enables the existence of companies like AirBnb and Uber, which are able to save approximately 30 per cent on labour costs by legally formulating their relationship with their workforce as "contractors" rather than employees who would require benefits, insurance, training, overtime, or sick pay (Srnicek, 2016, chapter 2). This business model is driven by the constant aim to expand into new markets, and according to Srnicek, is more about cutting costs than making money in the short term: bankrolled by constant venture capital inflows, the lean platform minimizes costs and the ownership of physical assets, while expanding globally and extracting the maximum amount of usable data (to hopefully help the firm to become profitable in the future).

Table 1.3: Political economy typology of platforms, based on Srnicek (2016)

Type	Examples	Key Feature
Ad Platforms	Alphabet, Facebook	provide services used to collect data for selling advertisements
Cloud Platforms	Amazon, Microsoft	rent out hardware and software
Lean Platforms	Uber, Airbnb	asset & cost minimization

Platforms and Political Salience

While this approach is more persuasive than any other existing typology, Srnicek does not fully flesh out his scheme or provide a complete classification of relevant companies. He suggests that some platforms matter more than others, mentioning that a company like Twitter is considered to be a "second-tier" platform (Srnicek, 2016, p. 102). It's an intuitive claim, but one which is provided without justification. Some platform companies surely have more political impact than others, but which ones? And how can we tell whether a platform matters politically in the first place?

First, there are the obvious quantitative factors, such as the size of the companies, both measured in financial terms as well as in terms of their user bases. For example, according to Statista (2021), Facebook has 2.2 billion monthly active users, compared to Twitter's 335 million; Twitter's net income of approximately 200 million USD is dwarfed by Facebook's net income of approximately 19 billion USD. Resources clearly make firms more politically salient, as they shape the ability of firms to wield traditional notions of corporate or business power (D. Fuchs, 2013): these include what Drezner, following Hirschmann, has called the usage of "political voice," including lobbying, political contributions, organizing public campaigns, and other instrumental forms of shaping legislation to one's interests (Drezner, 2008). For instance, in 2017 Google spent more on lobbying in Washington than any other company (Taplin, 2017), including traditional major lobbyists like Tobacco companies, Oil and Gas firms, and telecommunication providers.

Secondly, there are specific topical areas of significance, which vary across specific policy issue areas and across specific locales. For example, researchers looking at processes of political communication tend to look at firms like Facebook and Google: roughly 70 per cent of Americans use Facebook and YouTube, respectively, with the majority of those users relying on them every single day to access news and other types of political information (A. Smith and M. Anderson, 2018). Work looking at the cultural and social impact of platform companies on the every day lives of its users generally focuses on these two firms as a result (Noble, 2018; Vaidhyanathan, 2018). Perhaps the only equivalent to these two firms in terms of its daily political impact is Tencent, which has become an all-in-one space for political, social, and commercial activity in China through its WeChat app (Holmes, Balnaves, and Wang, 2015; Tu, 2016). Twitter is also a relevant player, despite its relatively small user base, the vast majority of whom are in the United States and the United Kingdom. While it is a niche platform for elites and journalists in most other countries, it remains a politically influential space for activists, civil society to organize social movements (Burgess and Baym, 2020; Freelon, McIlwain, and Clark, 2018).

Outside of the realm of political information, lean platforms such as Uber or Deliveroo clearly have the biggest implication for labour, local economies, and workers rights, with a growing body of scholarship documenting the strategies that employees, organizers, and regulators use as they struggle with the (often exploitative) rules and processes set by firms (Collier, Dubal, and C. L. Carter, 2018; Schor et al., 2020). Online marketplaces also can be politically salient: a company like Amazon can have significant global impacts, both in terms of affecting international supply chains, affecting local-level retail consumption patterns, and by structuring a market through its rules and policies. Work has documented Amazon's wide-reaching body of private law (Van Loo, 2016), and it has attracted the especial attention of competition regulators in the United States, the EU, India, and beyond (Khan, 2017; Khan and Vaheesan, 2017). In sum, platforms vary widely, and the platform companies that one focuses on are likely to also vary depending on the specific type of policy area one is interested in.

For questions of private rulemaking for online content standards, the most impactful are the advertising driven 'social media platforms' that, as they grew into larger and more profitable businesses, were forced by various interests — first economic, and then political — into creating what began in a relatively loose, ad-hoc manner, yet crystallized over-time into an extremely complex global ecosystem of private governance over public speech and behaviour (Klonick, 2017; Suzor, 2019). The topic has in recent years received growing attention from internet researchers, who have documented how the ecosystem of content moderation rule-making features not just policies set through a sprawling, private bureaucracy of rule-makers in Silicon Valley hashing out a global set of rules (Gillespie, 2018), but also their implementation through a global supply chain of contract labour (Roberts, 2019), with hundreds of thousands of workers (oft located in lower income Global South countries) tasked with evaluating content flagged for removal, buttressed by a growing number of complicated technical infrastructures and mathematical tools trying to do this work more efficiently, quickly, and cheaply (Gorwa, Binns, and Katzenbach, 2020). The following table illustrates the largest user-generated content based platforms, generally presumed to have the largest and most influential sets of private rule-making power over online discourse.

Table 1.4: Most widely used user-generated content platforms, Statista (2021). Note that Facebook now also owns WhatsApp and Instagram.

Platform	Users (millions)
Facebook	2740
YouTube	2291
WhatsApp	2000
Instagram	1221
WeChat	1213
TikTok	689
QQ	617
Douyin	600
Sina Weibo	511
Telegram	500
Snapchat	498
Kuaishou	481
Pinterest	442
Reddit	430
Twitter	353

1.2.2 How do Platforms Govern?

Thinking of platform companies as *companies* helps provide some analytical clarity as to the starting point from which to conceptualize their political effect and impact. There is a long tradition of thinking in political science about the problems posed by powerful private actors, with a wave of attention about the influence of corporations as political actors in global affairs beginning in the 1970s or so (Strange, 1991; Vernon, 1977). However, there has been very little work applying this lens to contemporary platforms.[3] On the flip side, there has been a large body of work in the interdisciplinary spaces of 'internet studies' and 'platform studies' — largely a combination of media studies, communication, digital sociology, law and technology, cultural studies, and organization scholars — that has looked at the various ramifications of platform services, but usually without anchoring them to broader debates about how corporate actors are expected to behave in world

[3]The notable exceptions from political scientists include the work of Culpepper and Thelen (2019) on Uber and Amazon, Haggart and C. I. Keller (2021) on political legitimacy and platform governance, and Tusikov (2019) on PayPal.

politics. These are two very broad bodies of literature that deal with the topic at hand, and I explore them here in turn.

Content Moderation: How (User-Generated Content) Platforms Govern

A long line of scholarship in the Human Computer Interaction subfield of computer science has studied how online communities create sets of rules, foster norms of common understanding and acceptable behaviour, and how they embody varying types of democratic cultures and political organization (Fiesler et al., 2018; Kraut and Resnick, 2012). Online forums and bulletin boards have since the 90s established mechanisms where community 'moderators,' generally volunteer members of that community, have the power to intervene in conversations, suspending or removing users, exerting significant labour behind the scenes to foster positive interactions and remove content that is in violation of community norms (Matias, 2019). Legal scholars observing these online communities in the early to mid 2000s implicitly adopted understandings of governance similar to those adopted by political scientists. For example, Grimmelmann argued that robust systems of community moderation and management as seen on bulletin boards, the early internet, and websites like Wikipedia are effectively "governance mechanisms" designed to "facilitate cooperation and prevent abuse" (Grimmelmann, 2015, p. 47). His taxonomy of 'content moderation' describes the varying ways in which governance in an online community works, with the community managers or creators deploying not only 'hard' interventions like content or user removal to curb harassment and set positive norms for acceptable participation, but also 'soft' architectures that structure activity — such as the use of reputation systems, ranking mechanisms, and the technical layout of a site (Grimmelmann, 2015). These architectures create structures that shape the behaviour of the users of online services.

The policy and design decisions that make up these moderation systems can either be made bottom-up by the community itself, or top-down, by the service owners and operators (Schoenebeck, Haimson, and Nakamura, 2020). In the 1990s and early 2000s, the status quo for early social networks — from bulletin

boards and USENET groups to fan forums and community groups — was volunteer moderation (Chandrasekharan et al., 2018; Kraut and Resnick, 2012). As they have become more commercialized and grown in size to have millions of users, sheer scale combined with a profit motive to lead today's major platforms away from an early reliance on volunteer moderators. While some large sites, like Reddit, with 430 million users, have retained the community moderation model, clustering into communities called 'sub-reddits' which have their own specific rules and volunteer moderators who evaluate user reports (Squirrell, 2019), the platforms that are large multinationals have largely adopted top-down models where the rules are created by the legal and 'product policy' teams in house, and then enforced through moderation enacted by a combination of internal employees and external contract labour (Caplan, 2018; Roberts, 2019; Wagner, 2013). To distinguish the dynamics of this type of moderation from its 'community' predecessor, digital media scholar Sarah Roberts coined the term 'commercial content moderation', or CCM, emphasizing it's industrial, outsourced and profit-driven nature (Roberts, 2018).

While scholarship on Commercial Content Moderation is still relatively nascent, with the field having only recently recieved its first book-length treatments (Gillespie, 2018; Roberts, 2019), work on moderation builds upon a long tradition of work in digital media studies that examines the implications of the technical, social, and algorithmic systems deployed by platform companies. "Critical algorithm studies" scholarship deploys insights from a wide range of disciplines to unpack the increasing role that automated decision-making plays in contemporary life (Barocas, Hood, and Ziewitz, 2013; Gillespie and Seaver, 2015; Ziewitz, 2016), necessarily implicating platforms. Scholars providing more specific explorations of how platform algorithms can encode bias at a cultural level (boyd and Crawford, 2012; Bozdag and Hoven, 2015; Noble, 2018), how users are constantly fighting to express their politics through their daily engagement with algorithmic systems (Bucher, 2017; Crawford and Gillespie, 2016), and how platforms mediate democratic participation and collective action (Milan, 2015; Tufekci, 2017) have all in effect delved into the varying ways that the services provided by major platform companies that are used

by billions of people around the world form part of today's contemporary political order. Content policies, terms of service, algorithms, interfaces — these are the governance mechanisms of online infrastructures (Plantin et al., 2018), and through their design and affordances, platforms can not only affect individual behaviour (Bucher and Helmond, 2018), but also have wide-ranging political impacts at a global level (Owen, 2015). Work from media scholars has further argued that digital platform corporations are increasingly shaping outcomes in the cultural and creative industries through a process of 'platformization' (Nieborg and Poell, 2018; Van Dijck, 2020). As Van Dijck, Poell, and Waal (2018) argue, the past several years have led to the creation of a 'platform society' which increasingly seeks to elevate privately developed norms, standards, and values over those of democratically constituted publics. As I have recently argued, this trend has sparked a significant amount of political contestation, with policy and academic conversations catalyzing around the question of how platforms govern their citizens — and how that governance should itself be governed and regulated by governments (Gorwa, 2019).

Governing Corporate Actors in International Politics

While international relations scholars have yet to comprehensively engage with this question of platform governance, they have much to bring to the table in terms of analyses of governance, private rule-making, and efforts to keep corporations accountable. The political concept of governance has evolved greatly in the past half-century. Initially associated primarily with domestic governments, governance was less a set of practices than a capacity: as per Fukuyama's traditional articulation, governance is the "government's ability to make and enforce rules, and to deliver services" (Fukuyama, 2013, p. 4). "Good governance," as commonly understood by political scientists, referred to a state's ability to build functional and effective institutions, and use those institutions to maintain law and order (Weiss, 2000). However, a movement in the 1990s towards 'global governance' in political science and international relations advocated a much broader understanding of governance (Rosenau and Czempiel, 1992). This more flexible notion engaged with the central

question of "how global life is organized, structured, and regulated" (Barnett and Duvall, 2004, p. 7), and sought to unpack the power relationships and conflicts that these structures could create or enforce. As Stoker put it, governance entails "creating the conditions for ordered rule and collective action" (Stoker, 1998, p. 17); it is thus more than just a capacity, but a specific and complex network of interactions spanning different actors and behaviours.

The post-WW2 international order, characterized by increasing interdependence through international organizations, institutions, and trade, enabled firms to expand globally, to the extent that transnational companies became "the most visible embodiment of globalization" (Ruggie, 2007, p. 821) during the 1990s and early 2000s. As the activity of firms became intertwined with virtually all important global social issues, from climate change and environmental damage to human rights and labour standards, corporations raised a vital set of governance puzzles (D. A. Fuchs, 2007; Hall and Biersteker, 2002). How could one get corporations to comply with human rights standards that were crafted specifically for states (Dingwerth and Pattberg, 2009)? How could firm behaviour be regulated across jurisdictions, and how could firms be nudged into more responsible business practices, often at the expense of their overall bottom line (Mikler, 2018)?

One of the major developing areas observed by scholars has been the emergence of private and informal forms of regulation. For instance, Abbott, Green, and Keohane (2016) have studied the emergence of private transnational regulatory organizations (PTROs), which are "established and governed by actors from civil society, business, and other sectors" and "engage directly in transnational governance, adopting standards of conduct for business and other targets on regulatory issues from worker rights to climate change; promoting, monitoring, and enforcing those standards; and conducting related administrative activities" (Abbott, Green, and Keohane, 2016, p. 248). As ample literature from regulatory politics and global governance scholars has documented (Abbott, Green, and Keohane, 2016; Büthe, 2010), a wide variety of private and informal governance has become increasingly common in both the domestic and international arenas, with governments increasingly

relying on informal arrangements to achieve certain aims (Roger, 2020), and non-governmental actors deploying various private standards setting arrangements for economic (signalling their commitment to policymakers and to publics)[4] or normative reasons (Renckens, 2020; Tusikov, 2017). The regulatory landscape thus includes not only traditional forms of regulation, but a host of voluntary arrangements, public-private partnerships, industry-specific measures, and many other collaborative arrangements with varying distributions of governance roles (Abbott and Snidal, 2009; Mattli and Woods, 2009; Zürn, 2018).

The regulatory scholar Natasha Tusikov has been a leader in analyzing the way that these private regulatory organizations play an influence in international technology policy issues (Tusikov, 2017, 2019). In *Chokepoints*, she traces the emergence of a largely informal, private regulatory regime developed at the behest of powerful economic interests (large multinational companies and brands) to combat the sale of counterfeit goods and other types of intellectual property violations online (Tusikov, 2016). Relatedly, Haggart (2014) has produced the most important international relations study of copyright politics, looking at the complex blend of interests, ideas, and institutions shaping online copyright regulation in the US, Canada, and Mexico.

The other group of scholars that have done the leading work that brings global governance in conversation with technology policy is based in Washington, DC. Farrell and Newman (2019) have undergone an ambitious agenda on the transnational regulatory politics of international data protection policies, publishing a book looking at how security focused actors in the European Union and the United States were able to built transnational regulatory networks and issue linkages that successfully resulted in changing data protection rules in both countries. Additionally, they have published a number of theoretically rich articles illustrating the core of a research agenda for what they call 'a new interdependence approach'

[4] As Abbott, Green, and Keohane (2016, p. 248) note, "voluntary standards rely on incentives such as consumer demand, reputational benefits, avoidance of mandatory regulation, and reduced transactions costs" but the varying types of actors involved in private regulatory efforts — ranging from civil society groups to firm associations — can be motivated by varying blends of normative and interest based factors.

that helps explain regulatory change in data protection rules, many of which implicate the services of the largest user-generated content platforms (Farrell and Newman, 2015, 2016, 2018).

1.2.3 Why Platforms Matter for International Relations

Some platform companies have, since their emergence in the mid-2000s, come to play a growing role in contemporary politics. Much political communication scholarship has focused on their use to coordinate social movements as a political campaigning strategy (Castells, 2012; Kreiss, 2012; Tufekci, 2017), but their role as political actors in global politics, however, has been less explored. While this may seem a niche topic for international relations scholars focused on interstate war, international organizations, and other classic fields of study, platform companies should be of interest to at least those sympathetic to global governance and transnational regulation issues for a number of reasons.

Firstly, platform governance should matter for international politics because of the actors involved. These are some of the most profitable corporate entities in history, and they arguably have developed the largest and most complex private regulatory orders that have ever existed. Secondly, these private systems of rule-making potentially have a bigger impact on the day to day lives of more individuals than most other forms of private governance. Whereas ordinary individuals may only have indirect interactions with companies when they purchase their products — a consumer might indirectly interact with a multinational like Exxon when they fill up their tank with petroleum products that it extracted — billions of people are at every minute interacting with interfaces created by companies like Facebook or Google, and these platforms in effect mediate their daily interactions with their family, friends, and public, especially in a pandemic era of working from home and social isolation. Platform governance can have a direct impact on the ability of people to communicate their identities, negotiate harassment and bullying, and live their lives free from domination, fear, or sexual exploitation, and thus has

high-stakes for individuals and their rights (Bivens, 2017; Duguay, Burgess, and Suzor, 2018; Matamoros-Fernández, 2017).

But platform governance also has important domestic and global political implications beyond the level of the individual. Platform content moderation decisions span multiple important policy arenas: while the creation of private rule-making, for instance, in an area like accounting standards, clearly implicates tax policy and some dimensions of international finance, platform-state relationships have the potential to have a tangible impact across a number of areas of political importance, from public health (e.g. by setting rules around vaccine information and misinformation) and immigration (e.g. by setting rules pertaining to the protection and denigration of refugees and immigrants) to national security (e.g. by setting rules that impact online radicalization and the circulation of extremist content). Additionally, platform governance poses some interesting geopolitical questions, with under-explored dynamics of technological development and international competition. Thus far, most of the most powerful platforms are American 'national champions' of sorts (with Europe bemoaning its lack of homegrown competition), but Chinese companies have not only emerged to play hugely important roles in Chinese social and political life (Holmes, Balnaves, and Wang, 2015; Plantin and Seta, 2019) but also begun to penetrate into the global market. On the horizon are not only potential clashes between Chinese and US-based systems of private regulation, but also broader strategic questions about how regulation designed to curb platform power might lead to the emergence of a Chinese-led system of global platform governance with significant economic, political, and cultural implications.[5] This overall has major implications for the current directions of globalization, and how increasingly economies are increasingly interdependent on rules and standards made in other jurisdictions (Farrell and Newman, 2019), and in this case, on the efforts of private actors to adjust their private rules with public impact to the interests and normative desires of various markets.

[5]This, at least, has been an increasingly popular argument advocated by American platform company CEOs when asked for reasons that their monopolistic practices should not be regulated. See e.g. Lyons (2020)

1.2.4 Gaps in the Current Research Landscape

Digital media, law and technology, and communications scholarship has blazed the trail when it comes to describing platform services and their systems of private governance. However, this work frequently misses the political dynamics underpinning these systems, and has generally under-appreciated the role of government and policymakers in shaping the ways in which platforms develop their regulatory orders (Gillespie et al., 2020).

On the IR side, there has been little effort in the past decade to channel the traditions of IR into the many technology policy issues that have arisen with mass digitalization (Kello, 2017), despite the clear applicability of global governance theories to a domain that features powerful multinational corporations and complex systems of private rulemaking. While early work from IR theorists applied global governance lenses to look at the regulatory regimes of the internet (Drezner, 2008; Mueller, Mathiason, and Klein, 2007), this work has for the most part not been updated significantly to explain what is a rapidly changing and evolving domain (The 'internet' of today looks little like what it did in the mid-2000s, given that it is now dominated by global corporations and is a far cry from the open, decentralized, governance commons that was the focus of scholars at the time; see Benkler, 2006).

Political science thus seems to be particularly well equipped to contribute to these important emerging debates, but has not done so. As mentioned earlier, there have been a few important exceptions to this lack of recent work: Tusikov (2016) and Haggart (2014) have deployed IR theories to assess the global politics of intellectual property, looking at how non-state interest groups (such as private companies, industry associations, and other rightsholders) have successfully mobilized governmental actors to implement rules for copyright and counterfeiting online through a mixture of domestic, transnational, and global regulatory regimes. Additionally, Farrell and Newman (2015, 2018, 2019) have undertaken a tremendously valuable research agenda on the transnational politics of data protection and surveillance, looking at the ways that officials in both the EU and US used transatlantic linkages of to successfully push their security-focused agendas despite frequent domestic opposition.

Relatedly, Newman and his collaborators have looked at key developments in data protection policy, like the European General Data Protection Regulation, looking at the role of corporate lobbying and other interest group preferences during its negotiation (Kalyanpur and Newman, 2019).

Despite this important and excellent work, there nevertheless remains a series of major lacunae in our understanding of platform governance. Firstly, while there has been important descriptive work from legal scholars seeking to depict the private regimes of governance created by major companies like Facebook and Google (Bloch-Wehba, 2019; Kettemann and W. Schulz, 2020; Klonick, 2017, 2020; Suzor, 2018), this work has done so in isolation from the formal and informal regulatory forces deployed by government actors that have sought to shape these practices. There is a clear need for a more holistic framework to understand platform governance more generally (Gorwa, 2019), and for a specific effort to link the regulation *by* platforms more clearly with the regulation *of* platform companies, to borrow the terminology used by Gillespie (2018).

Secondly, as legal scholars have observed the multitude of new regulatory proposals for online content and note that the domain is changing rapidly, we know little about the factors driving these changes, and whether established theories of regulatory change hold their explanatory power in this fast-moving regulatory domain which appears to have some unique issue characteristics. There have been no efforts yet, outside of narrow looks at copyright and intellectual property — which, I would argue, involve a different set of steering actors and economic incentives than the broader conversations about platform governance and online content regulation today — to apply the vocabulary and concepts of regulatory politics and global governance to the regulatory domain of online content regulation. If we expect change in platform standards to adhere to established theories of public-private regulatory contestation, there have been no efforts yet to test these theories.

Finally, there is an empirical gap in our understanding of key regulatory episodes and case studies, due to the lack of universe-level data and structured comparative analyses. While legal scholars have again published a number of focused textual

and normative assessments of varying regulatory frameworks and how they may adhere to, or conflict with, established practices of international human rights law or frameworks like the US First Amendment, there have been no systematic case studies of important regulatory episodes (such as the German NetzDG, the Australian AVM Act, the French Loi Avia, etc) to deploy a social scientific — rather than just purely descriptive and legal — method.

While this thesis will be unable to fill all of these gaps, I hope to at least provide a first step towards providing some insights that might contribute to our knowledge across all of these issue areas. This is a quickly growing interdisciplinary field, and I hope that this work will prove to be helpful in these early days as the issue of platform governance and platform regulation only continues to increase in importance.

2

Theorizing Contestation and Collaboration in Platform Governance

Contents

2.1 Introduction

The public and private regulative institutions of platform governance are shaped in a process of regulatory negotiation that involves multiple governance stakeholders — firms, various government actors, and civil society. But the specific patterns that this negotiation takes can vary significantly, with governments seeking to shape the private regimes of platform governance generally taking what can be broadly characterized as either a collaborative or combative approach to doing

so. Following the research question established in the introductory chapter, when governments seek to intervene and shape platform authority, under which conditions should we expect to get collaboration between platforms and governments, or when should we expect to see contestation from states seeking to 'take back control' in attempts to assert sovereign authority over politically salient transnational regimes of private platform governance?

This chapter develops a conceptual model of public-private authority in platform regulation, presenting a new typology of three strategies of platform governance: neutral, collaborative, and contested. These types describe the character of state-firm interaction in a specific governance constellation, and encompass different institutional configurations and varieties of private authority. *Neutral* platform governance occurs when a government either ignores, or tacitly accepts the existence of private platform authority over content moderation. Here, governments either ignore platforms, or seek to exert pressure on them via the established channels of complaint managed and set up by platform companies (e.g. by seeking to influence firm policy representatives and spokespersons). In this strategy, governments do not seek to fundamentally shape the ways in which platforms create transnational private regulatory orders that can have ramifications for a country's populace. In contested or collaborative platform governance, governments seek to shape the policies or practices of platform rule-making; either by collaborating with the companies to negotiate voluntary, non-binding rules, or by more fundamentally contesting the authority of companies by layering binding regulation domestically or transnationally.

Drawing upon global governance and transnational regulatory politics literature which seeks to explain the conditions under which changes in domestic or transnational regulatory frameworks governing corporate behaviour occur, the chapter presents a simple 'power plus' argument to explain variation between these three 'strategies' of platform governance. The argument is that the ability to engage in contested governance is a function of domestic and (in certain cases, transnational) power resources: if an actor seeks to contest platform authority, and has power

to do so via binding rules, the actor will do so. Governments that do not have the power resources to layer new rules via domestic legal frameworks, and yet still seek to shape platform's content moderation practices, will need to resort to (less costly, but also less enforceable) voluntary, collaborative mechanisms. However, I argue, the decision to engage in collaborative versus contested governance is not purely explained by power — it depends on the character of the demand for new rules, which, following historical institutionalist insights, is heavily shaped by the domestic institutional context (in particular, normative understandings of the acceptable degree of government intervention in the area of freedom of expression, and institutional interdependencies that may constrain action). For this reason, under some conditions, state actors with the power resources required to contest private governance via binding rules will instead choose to pursue collaborative arrangements.

In the first section, I present my typology of platform collaboration/contestation, building upon existing work on private authority and private rulemaking in global politics. The second section provides an overview of key concepts from the international regulatory politics literature, and uses these building blocks to articulate the main theoretical argument of the thesis.

2.2 A Typology of Platform Governance

Theories of Private Authority

Platform companies are, of course, not the only private actors that have created politically salient systems of private rules. International relations and global governance scholars have, from at least the 1980s onwards, recognized the potential influence of non-governmental organizations and civil society in creating guidelines, standards, and rules that could shape corporate and governmental behaviour (Keck and Sikkink, 2014; Tarrow, 2005). In the past decades, there have been codes of conduct, 'Global Compacts', corporate social responsibility creeds, and a host of various multi-lateral, international, and domestic level governance initiatives spurred by political pressure from states and civil society, ranging in their impact, complexity,

and institutional configurations (Bernhagen and Mitchell, 2010; Ruggie, 2007; Sikkink, 1986). Similarly, firms in a multitude of industries have themselves — either in isolation, or in partnership with government and/or civil society stakeholders — developed various transnational regulatory schemes that can take many forms (Bulkeley et al., 2012), ranging from industry associations that enact self-regulatory codes to broader initiatives that bring together a host of different actors, and often include participation from civil society or government (Fransen, 2012). The result has led to a huge number of private governance initiatives have been developed by groups of NGOs and industry groups, with dozens of efforts that have sought to create standards and outline best practices around sustainability (e.g. ISO14001, the Forest Stewardship Council), labour rights (e.g the Fair Labour Association, the Workers Rights Consortium) and many other areas (Büthe and Mattli, 2011; Dingwerth and Pattberg, 2009; Fransen and Kolk, 2007). A recent effort to collect and categorize data on as many of these private transnational governance initiatives as possible found close to 700 initiatives across more than 30 policy issue areas ranging from tourism to technical standards (Westerwinter, 2021).

The core novelty in these types of private, or public-private regulatory arrangements, according to Abbott and Snidal (2009, p. 505), is the "central role of private actors [...] and the correspondingly modest and largely indirect role of 'the state,' " as well as the fact that "most of these arrangements are governed by firms and industry groups whose own practices or those of supplier firms are the targets of regulation." Importantly, these types of arrangements, even when informal, still involve power, featuring differing balances of interests from different actor groups (Raymond and DeNardis, 2015). Power relations are not only at play during the negotiation of new rules, practices, and other regulative institutions across an initiative's constituent stakeholders, but also in terms of outcomes — whether the initiative becomes powerful in terms of having an effect on other actors and issues.

Most of the scholars working on this type of private rule-making power in global politics focus on *authority*, rather than power. Drawing upon the fundamental work of political theorists like Hobbes and Weber, IR scholarship today frequently deploys

a definition of power that focuses on coercion, drawing on the idea that if one actor is able to force another to do something, the first actor holds power over the other (Barnett and Duvall, 2005; Dahl, 1957). As Green (2014, p. 43) notes, "Unlike power, which may involve coercion, authority implies that consent is granted by those who are subject to it... authority is a type of power and can be understood as the right to command." Rule-making by either tacit or clearly delegated consent has become a major object of study in global governance and transnational regulatory politics, spanning a wide-range of the aforementioned policy areas, such as global financial regulation, environmental management, and technical internet protocols (Büthe and Mattli, 2011; Cashore, Auld, and Newsom, 2004; DeNardis, 2009).

In an early and influential edited volume on private authority in global governance (Hall and Biersteker, 2002), the collected essays argue for three broad types of authority exerted by private actors: market, moral, and illicit. In another landmark collection, Avant, Finnemore, and Sell (2010) argue that private governors can become powerful due to their institutional, delegated, expert, principled, or capacity based authority. There are many overlaps between these concepts, but the core is that authority is derived either from the impact private actors have on other actors — for example, private financial actors can make decisions that have market-based ramifications for national institutions and legal frameworks, thus yielding private market authority (Hall and Biersteker, 2002, p. 12) — or from the expertise and legitimacy that the private actor has, leading it to either be able to shape future outcomes, or perhaps to have governance functions delegated to it (Avant, Finnemore, and Sell, 2010, p. 2). To borrow some legal terminology, a distinction might be made between either *ex ante* or *ex post* private authority; private authority that is gained by the private actor leading, or by the private actor being followed by others. For example, non-state armed groups operating in areas of limited statehood that make decisions that have major impacts on the people living in those areas are *de facto* exerting one form of ('illicit') private authority (Börzel and Risse, 2010); non-governmental organizations that create international standards that are adopted by others due to the reputation and expertise of that organization

are exerting a slightly different brand of moral/expert private authority (Green and Auld, 2017; Mattli and Woods, 2009).

Green (2014) seeks to provide a more concrete empirical framework for understanding these various dimensions of private authority in world politics. Her understanding of private authority is more narrowly concerned with *ex post* authority, defining it "as the ability of non-state actors to make rules or set standards that other relevant actors in world politics adopt" (Green, 2014, p. 46). Helpfully, Green's typology involves two forms of private authority, 'delegated private authority' and 'entrepreneurial private authority.' In Green's account, the difference between delegated and entrepreneurial private authority hinges on whether private actors have received permission from the onset to act on behalf of other (state) actors; if not, they must persuade others to adopt their rules by 'entrepreneurially' building legitimacy and credibility (Green, 2014, p. 61). For example, if a group of governments delegate the upkeep of technically complex international accounting standards to a private organization largely made up of experts from the private sector, that is a marriage of convenience through which the private organization has in effect been 'granted' (potentially wide-ranging) authority in a policy area. On the other hand, if a group of firms get together to create an industry-wide certification scheme for paper products that fit some measure of sustainable production, that is an informal, 'entrepreneurial' regulatory arrangement that then could be voluntarily adopted by other firms, or even recognized by states in formal regulatory arrangements (e.g. passing domestic laws featuring incentives to firms to join the certification scheme) and thus have a authoritative impact.

2.2.1 Neutral Platform Governance

The work of Green and other IR theorists working in the space of private rule-making in transnational regulation is instructive for the questions I am seeking to answer, but not conclusive. Firstly, Green's model seeks to explain the conditions under which private authority might emerge, but does less to explain patterns of change over time, and the conditions under which it may clash with government authority,

and be 'un-delegated,' contested, or withdrawn. Secondly, existing work has been extremely helpful in providing explanations for why private rule-making occurs, but it does less to comprehensively map out the relationships between different types of actors that lead to specific types of outcomes and governance configurations. Because the aim of this thesis is *not* primarily to understand the various dimensions of private authority, but rather to address the varying relationships of governance between different actors in public/private platform governance constellations, the theory needs to be taken a step further.

Based upon the analysis of formal and informal regulatory arrangements presented in Chapter 3, a few empirical observations about the historical evolution of platform governance can be made. First, the rise of wide-scale, impactful private rulemaking for online content developed in a manner akin to the patterns of entrepreneurial authority described by Green (2014). Today's (relatively speaking) old user-generated content platforms, like Facebook and YouTube, all developed their content moderation processes in an incremental, ad-hoc fashion, largely due to an awareness that their failure to do so would drive away advertisers and harm their commercial interests (Gillespie, 2018). They were created in a specific institutional context created by laissez-faire baseline intermediary liability laws like the US Communications Decency Act and the European Union's E-Commerce Directive, which allowed platforms to make commercially-driven decisions about online content without taking over legal responsibility for that content (Klonick, 2017); however, the specific modern function of major platforms as de-facto custodians of the digital public sphere was never specifically delegated to them by willing governments. Instead, a relatively marginal piece of legislation, crafted for a still relatively marginal industry (in the US Case, online bulletin board services like CompuServe and America Online) would provide a playing field upon which platforms were able to — as Gillespie (2018) has argued, begrudgingly — create increasingly complex rules over online speech and action (Kosseff, 2019).

This context, which I call *neutral platform governance*, is the governance status quo for much of the world. In this mode, governments do not fundamentally oppose

the systems of private rule-making over online content created by the platform companies operating in their jurisdictions, and nor do they seek to exert significant political capital or power resources in an effort to shape those systems. Online content moderation may be a low salience issue, where it simply does not matter much to those in power; or, for various other reasons, the status quo is satisfactory enough that policymakers do not seek to overhaul the foundational, laissez-faire institutional frameworks upon which online content moderation relies. In effect, platforms exhibit entrepreneurial authority,[1] and the source of this authority is either tacitly tolerated or ignored by governments. While governments may seek to obtain certain governance *vis a vis* platforms, in this strategy, they do so within the established practices of corporate-government engagement — for example, by sending takedown requests for certain pieces of content, or requests for user data, to local firm representatives (Schwemer, 2019). The relations between firm and government, despite the 'neutral' moniker, may be quite strained, and the government may exert various levers of pressure against the firm to takedown pages, groups, or individual pieces of content (Kaye, 2019); nevertheless, this occurs within established channels and does not manifest in broader efforts to fundamentally change how platforms govern content.

2.2.2 Collaborative and Contested Platform Governance

When a government seeks to shape firm-led regimes of content moderation in some way, it moves away from the 'neutral' state and towards a space where it is a more active participant in shaping private platform rule-making and rule-implementation authority. The first point to note is that platform rule-making is

[1]Importantly, this is not exactly the kind of entrepreneurial authority defined by Green, who bases her definition of authority primarily around the adoption of specific rules willingly by other actors — with the idea that authority exists when a standard, guideline, or rule (such as a voluntary certification code) made by actor X is willingly adopted by actor Y. Instead, this is a broader idea, where platform rules do not need to be adopted or implemented by other platforms — but rather, the rules are implicitly used or tolerated by large numbers of users, affecting their lives directly or indirectly. I would argue that Facebook can gain private authority when its Community Standards are tolerated by the billions of customers that have accepted the Terms of Service to begin using the service, and not just when YouTube, for example, were to copy a process or specific ruleset from Facebook.

largely regulated through what Farrell and Newman (2010, p. 11) call 'international market regulation': "the processes through which the domestic regulatory activities of states and other actors set the effective rules of internationally-exposed markets." In the absence of major international fora for digital policy decision-making at the content or application layer of the internet, platform rules are shaped by domestic regulations in various jurisdictions, as well as a smaller number of informal and voluntary transnational governance initiatives (see Chapter 3). Therefore, the process of collaborative or contested platform governance primarily originates at the domestic level.

Empirically, we can observe two substantive differences in the type of regulatory frameworks that government actors seek to deploy to shape platform rule-making: whether they do so via a formal, or informal mechanisms. Do states seek to contest platform rulemaking by instituting formal legal frameworks with binding commitments (that have legal repercussions, and enforcement)? Or do they seek to negotiate changes via the public-private negotiation of voluntary commitments?

Depending on whether the state pursues the former, or the latter, the outcome is either *contested platform governance* or *collaborative platform governance*. In Table 1.1, I outlined the two substantive institutional dimensions that characterized these two strategies.

Substantively, the key elements of institutional variation are the (a) formality of the commitments made in the regulatory initiative, and (b) the role of the actors involved — more specifically, which actor is supplying the newly negotiated rules. Following the observation made by Büthe (2010) that there are three main stylized roles involved in regulatory politics — the demanders, suppliers, and targets of rules, and that different mixes of these roles underpin different types of regulation[2] — a key difference lies within whether the new rules being implemented are were designed by a state actor, or if they are supplied by firms or a combination of different actors. Were new regulations developed just by government(s), possibly

[2]For example, in the ideal-type of pure self-regulation, a corporate actor is the demander, supplier, and target of a regulatory initiative; in 'command and control' regulation, the state is the demander and supplier, and the firm the target (Büthe, 2010, p. 8).

with some sort of stakeholder consultation or influence (e.g. lobbying), or were they developed in what might be more accurately called a multistakeholder process?[3]

Additionally, there is a discursive, rhetorical characteristic that can be empirically observed, which often, but not always correspond to these particular governance strategies. Firstly, a collaborative platform governance style involves rhetoric that emphasizes the importance of cooperation, working together, and building capacity via partnership. The language is that of partnership, and sometimes, even equal partnership, between governments and platforms engaged in a specific governance enterprise. In contrast, a contested, combative governance style instead features a rhetoric of control — a rhetoric that emphasizes the illegitimate nature of private rules, and seeks to exert state primacy over them, either through a discourse of sovereignty or securitization. It also features more of a coercive element: the regulatory targets are rhetorically subject to threats, 'sticks' rather than 'carrots,' with the logic being more one of power and sanctions rather than coalition building, cooperation, and voluntary commitments. However, the substantive/institutional element and the stylistic/discursive element are usually, but not necessarily, aligned. It is possible for a government to engage in a substantively collaborative governance effort, one which yields voluntary, public-private commitments, that is still underpinned by contested rhetoric, such as a rhetoric of sovereignty and the threat of future legislation in the case of non-compliance (as seen in the case of the EU Codes of Conduct around Illegal Online Hatespeech, for instance; see Citron, 2017). Inversely, it is possible that a government passes domestic binding rules that contest the ability of a platform's private rule-making systems, and yet does so while professing a rhetoric of partnership and

[3]Multistakeholder governance, has been neatly defined by Raymond and DeNardis (2015, p. 573) "as two or more classes of actors engaged in a common governance enterprise concerning issues they regard as public in nature, and characterized by polyarchic authority relations constituted by procedural rules." In other words, it is governance involving actors from at least two of four groups — states, nongovernmental organizations (including civil society, researchers, and other parties), firms, and international organizations like the United Nations — where decision making authority is distributed in a 'polyarchic' or 'polycentric' arrangement where one actor does not make decisions unilaterally (Black, 2008). Here, I understand rules supplied by a 'multistakeholder' group to involve at least two of the three broad actor groups involved in platform governance: industry, government, and civil society.

collaboration with technology companies. For this reason, the key variables that determine whether the governance mode is collaborative or contested come down to the institutional character of the rules that are being deployed in an effort to shape the policies and practices of platform content moderation.

2.3 Explaining Variation in Platform Governance

Having outlined a simple typology with three potential forms of government-led change in platform governance, the next step is to advance an argument as to when, and under which conditions, these different strategies of pursuing change platform governance should emerge. In effect, this is a question of regulatory change, and international relations, international political economy, and regulatory politics scholarship has highlighted a number of potential causal mechanisms and explanatory factors to consider.

Perhaps the simplest, most elegant argument is advanced by Drezner (2008), who argues that it comes down to market size and state power: if large, powerful states desire a change to the status quo, then they (a) usually have the regulatory toolbox to leverage coercive force against firms to gain compliance, and (b) the targets of regulation are additionally more likely to bear the costs of complying with the new rules in order to maintain access to a large and profitable market. Following this logic, we would expect large, powerful states (Drezner refers to them as the 'great powers,' like the United States, European Union, and China) to be willing and able to contest platform private authority via regulation.

Drezner's model is primarily intended to explain international regulatory regimes — in other words, international outcomes — but it does less to explain "domestic outcomes that have international repercussions" (Farrell and Newman, 2010, p. 611). Because platform regulation empirically begins within a jurisdiction at the domestic level, there are three key variables that I argue shape whether a state is able to successfully change the platform governance status quo in their jurisdiction: demand for new rules, the power to intervene, and whether the rule-makers are influenced by either what I call a 'norm of intervention' or a 'laissez-faire' norm.

2.3.1 Demand Factors

In more legally-oriented studies of regulation and regulatory politics, the central driver of regulation is *demand* for regulatory change from policymakers, which should occur when knowledge of market failures, such as negative externalities, coordination problems, information inadequacies, or other harms arises (Baldwin, Cave, and Lodge, 2012, p. 15). In other words, regulatory change requires the preferences of policymakers to be aligned with that change. Much scholarship in political behaviour has sought to examine the sources of policymaker preferences, creating models based upon assumptions of what constitutes rational behaviour for policymakers, such as the motivating desire to be re-elected (Mayhew, 2004), to build influence, and to achieve policy goals (Fujimura, 2016). At an aggregate level, more 'liberal' varieties of rationalism in international political economy conceptualize preferences as based on "shifting pressure from domestic social groups," where those "preferences are aggregated through political institutions" (Moravcsik, 1993, p. 481).

These preferences are not only material; even hardcore rationalists are increasingly noting that actors do not simply seek to fulfill their material interests. As Abbott and Snidal (2000) have argued, studies of international rule-making require a theoretical framework that can account for interest-based explanations of demand well as normative ones. As actors use rules to "achieve their ends whether they are pursuing interests or values," and because "rules and institutions operate both by changing material incentives and modifying understandings, standards of behaviour, and identities" (Abbott and Snidal, 2000, p. 425), actor interests, and thus the demand for regulatory change, can potentially be affected by ideas (e.g. shifting public opinion, risk perception, and other discourses), values (such as human rights), and cultural norms or roles.

Demand is actively a site of contestation, with a plethora of interest groups seeking to affect the preferences of the 'demanders' for new rules when the stakes are high. By lobbying and deploying various forms of structural business power, firms can seek to dampen the demand of key decisionmakers for change, for example, by providing financial contributions to re-election campaigns or threatening to 'exit'

and take investment and jobs out of the country (Mikler, 2018). In certain cases, government rule-makers can even be 'captured' by business interests, so that they either do not demand changes at all, or if they do, demand firm friendly policies (E. Keller, 2018). Civil society groups and transnational advocacy networks also often seek to affect policymaker demand through various strategies, including by engaging in their own policymaker-focused lobbying, advocacy, and expert consultation, as well as by building public relations campaigns and other 'grassroots' initiatives seeking to mobilize constituents to pressure their representatives to intervene on their behalf (Keck and Sikkink, 2014; Tarrow, 2005). Platform firms, in particular, have countered by leveraging their direct access to consumers (via the devices and apps in their pockets) in an effort to try and mobilize them *against* proposed regulations, frequently arguing directly to their users that rule changes may increase prices or otherwise hamper their favourite services (Culpepper and Thelen, 2019).

States can also seek to affect the demand of other states for domestic regulatory change, and lobby or exert diplomatic influence in an effort to depress demand in other countries. This can occur when new regulatory changes are perceived by a powerful state as against its interest: for instance, as part of the US State Department's 'Internet Freedom Agenda,' the US government in the mid-2000s to early 2010s actively sought to maintain a minimal regulatory environment for internet-related services as part of a broader political, economic, and foreign policy agenda (Powers and Jablonski, 2015). It can also occur if firms are able to successfully lobby their 'home' state to oppose those foreign regulations on their behalf: for example, as Bradford (2020) describes, American chemicals and manufacturing firms galvanized the US government to lobby hard against complex EU chemicals regulation in the mid-2000s, with the US exerting economic and diplomatic pressure not only in Brussels, but also via American embassies and consulates in individual member states as part of a broader effort to minimize costs to American firms and maintain US economic competitiveness in the area.

These broad, macro-level notions of demand are affected by issue-area specific dimensions which help determine how actors in different jurisdictions may demand

different forms of rules affecting the online environment more broadly and the services created and managed by platform companies more specifically. Policymakers may wish to protect their constituents from content that can be harmful to either public safety or public health (e.g. calls to violence, misinformation about vaccinations), or that harms individual rights and freedoms (e.g. hate speech, child abuse imagery), but the norms around the extent to which these different issues are understood to be of national importance vary across communities and across jurisdictions. Different stakeholders engaged in regulatory contestation over the boundaries of acceptable content online might demand more or less government and firm intervention depending on their preferences; for instance, child safety NGOs are likely to demand higher standards than firms or NGOs dedicated to protecting civil liberties and free expression. These various sources of demand will vary significantly across countries and contexts, but will combine to shape government decision-making as policymakers decide to eventually lean one way or another and demand change.

2.3.2 Supply Factors

The Power to Intervene

Once demand for change is at a sufficient level — outweighing the resistance of other interest groups or stakeholders — political actors also to be able to meaningfully *supply* those changes for change to occur. Following 'supply and demand' models of public-private business regulation (Büthe, 2010), demand for changing the regulatory status quo is a necessary, but alone insufficient condition for change in platform governance configurations.

For rationalist IPE scholars, the classic explanation of supply comes down to state capacity and power, with market size usually deployed as a proxy for government power (Drezner, 2008). The logic is that large and powerful economies are likely to have larger regulators, more resources, and more expertise, and furthermore, big markets matter more for the firms that might lose access to those markets, so they are more likely to comply with government regulations in those jurisdictions (Newman and Bach, 2004). In these power-driven theories of international political

economy, countries with the largest economies in effect are able to intervene in transnational regulation as they wish, dictating the (international) rules that suit their interests, with regulatory fragmentation occurring when the most powerful states have clashing preferences (Drezner, 2008).

Nevertheless, other, more institutionally minded scholars have sought for at least a partial rebuttal of this power-centric argument, arguing that shaping firm behaviour is inherently difficult, even for the most powerful actors. Even in a 'traditional' or conventional regulatory relationship where a state actor demands and supplies new rules to bind a corporate target, the firm's managers and internal structures will at the end of the day be required to implement those rules (Abbott and Snidal, 2009). More complex 'decentred' or 'polycentric' regulatory regimes can additionally involve a host of regulatory intermediaries, such as third parties involved in the monitoring, implementation, or enforcement, of new rules, as well as other actors (Abbott, Levi-Faur, and Snidal, 2017; Black, 2008). For this reason, designing thoughtful and effective regulation for private actors — and then being able to meaningfully enforce that regulation or sanction non-compliance — requires significant regulatory capacities (Bradford, 2012; Saurwein, 2011). Because of this distribution of roles and competencies, "the ability to define, defend, monitor, and enforce a particular rule-set" is also essential for government actors seeking to make the kinds of credible threats needed for firms to take domestic regulation seriously (Newman, 2017, p. 82-83; Bach and Newman, 2010). This is especially important in potentially complex technology policy domains: as Saurwein (2011, p. 344) puts it, regulatory capacity has a major impact on the ability of an actor to intervene decisively and set rules for the targets of regulation on a policy issue, with the availability of adequate means to adopt and enforce statutory regulation determining whether a sub-state policy actor can actually make credible commitments and threats. In the domestic context, this involves power balances and rule-making configurations — does the rule-making branch of government have the ability to unilaterally create new rules, either due to an authoritarian grip or due to a commanding electoral majority, or does it need to make difficult coalitions across parties? These domestic

factors all matter for an actor's power to intervene when there is demand for new rules, and while the notion of the ability to supply rules is often correlated with market power and state size, there are exceptions — smaller states with strong regulatory capacities and highly competent regulatory agencies exist as well.

Institutional Constraints

The power to intervene is also shaped by domestic and transnational institutions and the constraints they impose upon rule-makers. Although the most rationalist scholarship assumes that the status quo is highly malleable, and can be changed whenever demand from powerful actors is sufficiently high (Drezner, 2008), on the other hand, scholars more aligned with historical institutionalism (HI) see institutions as more durable and less malleable than rationalists (Jupille, Mattli, and Snidal, 2017). Going beyond just power and interests, historical institutionalists emphasize the importance of regulative, normative, and even cognitive structures, noting that political life is structured not only by "formal and formal rules, monitoring and enforcement mechanisms, [but also] systems of meaning that define the context within which individuals, corporations, [...], nation-states, and other organizations operate and interact with each other" (Campbell, 2004, p. 1). HI scholars see institutional outcomes as reflective of hard-fought political battles and process of political contestation, leading to their interpretation that institutions are deeply embedded into political systems: as Pierson (2000, p. 262) puts it, "the key features of political life — public policies and (especially) formal institutions— are change resistant [...] Both are generally designed to be difficult to overturn." The corollary is that decisions made about certain institutional configurations can have potentially significant and possibly unforeseen effects on the institutional 'path' and the types of choices that are available at future political junctures (Thelen and Steinmo, 1992).

For our purposes, a number of institutional dimensions are especially important. Domestically, pre-existing institutional structures shape the configuration of regulatory agencies and government bodies that are potentially equipped to handle issues

of online content that straddle media, telecommunications, and internet policy domains. They also include the process of making, implementing, and enforcing policy, including the various 'veto points' "where the mobilization of opposition can thwart policy innovation" or change (Thelen and Steinmo, 1992, p. 7) that vary across political systems. Transnationally, they include broader relationships of complex interdependence characteristic of 21st century global politics, which include a host of economic, diplomatic, and political linkages with other states (Farrell and Newman, 2014, 2016). These institutional interdependencies might include formal institutional agreements scoping the range of policy change available — for instance, commitments made as part of a regional bloc, such as European requirements that member state legislation comply with existing European legal frameworks, or concessions made in a trade agreement that a country will not change its (intermediary liability or other) rules (Krishnamurthy and Fjeld, 2020). These interdependencies may also include informal, individual-level relationships of influence, for example between like-minded policy officials engaged in transnational regulatory networks (Farrell and Newman, 2019; Slaughter, 2004), or state-led mechanisms of international economic coercion via threats and sanctions imposed *vis a vis* individuals, firms, or governments (Farrell and Newman, 2019). All of these potentially important factors may conceivably have an impact and potential constraint on a government's power to intervene in various contexts. Taking this in combination with the understanding of power and regulatory capacity advanced above, on the supply side of this argument, I understand the power to intervene simply as: *regulatory capacity - institutional constraints = power to intervene.*

Normative Landscape

Historical institutionalist lenses lend themselves well to complex patterns of political change, including those potentially not just explained by actor agency and power during a specific period, but more intangible, contingent factors like identity, beliefs, and norms. Systems of meaning, often summarized as 'ideas', are especially important to historical institutionalists (Majone, 1998; Weir, 1992). Historical

institutionalist scholars are interested in the political role and evolution of 'macro-level' ideational structures (like class, or capitalism), and are also especially interested in the ways in which the preferences of actors are themselves shaped by various norms or ideas (Fioretos, 2011). As Thelen and Steinmo (1992, p. 8, emphasis in the original) note, in contrast to rationalist approaches, historical institutionalists assume that "not just the *strategies* but also the *goals* actors pursue are shaped by the institutional context." In other words, norms, ideas about appropriate behaviour, and pre-existing world views, themselves shaped by macro-level institutions and individual historical and cultural contexts, can be expected to influence the preferences and behaviour of actors, and to shape their ability and desire to supply certain types of policy change. History, culture, and overarching value-based frameworks about how regulation should work, and what the appropriate strategies for regulating technology, the media, and communications are, vary significantly across jurisdictions and also can impact the type of platform regulation that is pursued in various contexts.

The challenge for empirical analyses in this area is that these important factors are difficult to unpack and operationalize without deep historical work and thick, single-case study analyses that could command theses or books of their own. Nevertheless, one way to factor in normative factors into this model is by thinking of it as an additional supply factor, working alongside a state's power resources. One way to conceive of this factor is through employing some features of the constructivist concept of a 'logic of appropriateness' (March and Olsen, 2011), which has been shown to be important in regulatory politics given the oft well-defined roles and identities of certain governmental actors, especially regulators, civil servants, and judges (Eberlein and Radaelli, 2010). As actors have an understanding of what their acceptable scope of action is, and even if they may be driven by the interests of their constituent groups to desire certain changes to the status quo, they can be constrained by the socialized understanding of their appropriate intervention. Naturally, norms also affect the demand side, and the key issue areas that government actors may see it as in their interest to protect. But they also may

affect their willingness to intervene and achieve those normative goals. For example, two countries might have a commonly held normative understanding regarding the importance of prohibiting and combatting racist speech, but one country might see it as the role of the state to limit such speech, and the other might see it less so, and perhaps the role of firms or other actors.

In this case, what I call the *appropriateness* that an actor sees towards supplying certain types of rules relates to their internal perception of their own mandate and role. I argue that, in platform governance contexts, there are two broad sets of normative logics which are important to explore: the prevailing attitudes to corporate regulation broadly and technology regulation more specifically amongst prospective rulemakers (which may be regulatory, or de-regulatory), and their attitudes towards free expression and free speech (which may be absolutist, or constrained). Together, these factors create what I call either a 'laissez-faire' or 'interventionist' normative context.

Laissez-Faire Norm

Since the 'neoliberal' ascendancy of the second half of the 20th century, a number of the world's major economies, especially the United States and the nations of the British Commonwealth, experienced a turn towards 'de-regulation' and decreased state involvement in key industries (Derthick and Quirk, 1985; Hammond and Knott, 1988), in contrast to more corporatist, mercantilist economies like Germany and Japan that maintained a closer relationship between powerful national champions and the government (Cortell and Davis, 2005). In the United States, early internet infrastructure was formed in a crucible with defence and government funding and a mixture of countercultural and libertarian ideologies, creating the early context for the emergence of internet companies like Google and Microsoft (Abbate, 2000; Turner, 2010). Bietti (2021, p. 4), in a historical overview of the concept of digital platform regulation, argues that in "the digital platform context," the Californian strand of "anarcho-libertarianism has retreated and morphed into a libertarian aversion to regulation as well as a series of market optimist perspectives"

that have historically shaped debates around platform regulation in the US and continue to do so to this day.

The regulatory context in the United States, important as the world's largest economy and the home state for almost all politically powerful global platform companies, is shaped by the particular American tradition of free speech, grounded within a First Amendment that some legal scholars have argued is inherently deregulatory in nature due to its protections for commercial speech (Garden, 2016). In this context, a powerful norm against close government control over vehicles of free expression can exist (Kosseff, 2019), as seen not only in press freedoms and laws, broadcasting regulations, but also a general hesitancy for government to wade too closely into the muddy waters of shaping specific patterns of content distribution. In surveys of free expression norms, including codified conceptions of free expression in constitutional documents, countries like the United States consistently rank as having a more permissive, absolutist environment for free expression, with very few criminal and civil limits over individual speech (Krotoszynski, 2006).

The combination of general regulatory attitudes around communication infrastructure and attitudes around free expression combine to create a *laissez-faire* normative environment which may constrain policy change in certain contexts. While the United States is a clear example of this, conceptually, one might conceive of contexts, where rulemakers, while not necessarily absolutist around free expression speech, have an ingrained normative appreciation for the potential pitfalls — in terms of perceived excesses in government control, human rights harms, or other freedom of expression issues posed by platform regulation — that lends itself to a hesitancy to directly intervene in this complex and politicized regulatory environment. When enough important government stakeholders exhibit these attitudes, they can be said to harbour what I call a *laissez-faire norm* around regulating online expression.

Interventionist Norm

Of course, norms about how information services should be regulated vary greatly around with the world, with, for example, Germany, the United States, China,

and Russia, all having different notions of the appropriate level of government involvement in the online ecosystem and the extent to which it is acceptable for regulators and policymakers to get involved in regulating and limiting expression more generally (Fukuyama and Grotto, 2020). Many countries around the world have a considerable amount of government influence over media policy — funding public broadcasters, offering media stipends, or having oversight authorities of various forms that ensure the application of legal principles or codes of conduct that bind information distribution organizations (broadcasters, telecommunications providers), and there has been a large body of work in comparative media systems research seeking to document and understand these various policy characteristics (Hallin and Mancini, 2004; Puppis, 2010). In more authoritarian or single-party systems without free and open elections, domestic information distribution channels may be directly under state control, and are seen by policymakers as clearly part of their sovereign governance mandate. In China, for instance, policymakers have historically seen the "information sphere" as a clear extension of their offline jurisdiction, authority, and sovereignty (Plantin and Seta, 2019). In the hybrid regimes and monarchies/theocracies of the Gulf, a wide-range of information controls, and narratives around security, sovereignty, and cybercrime have also steadily sought to bring the online domain under the normative banner of rightful state intervention (Hassib and Shires, 2021).

Additionally, the global landscape of free expression norms varies significantly, as outlined in various comparative legal analyses of free speech policies, and the specific values underpinning them (Carmi, 2008; Krotoszynski, 2006). While free expression is famously part of Article 19 of the Universal Declaration of Human Rights, that principle is not unconditional, allowing restrictions for a wide range of reasons including the protection of the rights of others, national security, public order, and more (Benedek and Kettemann, 2014). Most countries have at least some restrictions on the acceptable scope of free expression, and some countries, including many of the world's biggest economies, have many. In Germany, as Tworek (2021) outlines, a 19th century Prussian and Imperial German legal tradition of

speech restrictions combined with the post-World War West German efforts at de-Nazification to yield a unique framework where dozens of different types of content — ranging, famously, from the symbols of banned organizations such as the Nazi Party to material that seriously defames religious or ethnic groups — were made illegal under the German criminal code. Similar provisions for criminal sanctions against those engaging in discriminatory, hateful, or racist speech exist not just in many European countries, but also in numerous Latin American countries, including Argentina, Bolivia, Brazil, Cuba, Guatemala, Mexico, and others (Hernández, 2010). These normative frameworks come from a different ideological tradition than free-speech-absolutist contexts, and are grounded less within a notion of individual liberty and freedom, and more within an idea of individual dignity and the need to protect it (Carmi, 2008).

The combination of a more permissive attitude towards regulating the means of political and public information distribution, as well as a more restricted notion of politically appropriate expression, can create an *interventionist norm* which can impact the ability of policymakers to supply new types of policy change in areas that directly affect free expression. In effect, I posit that in countries with a tradition of regulating across various content related information sectors, and with normative traditions of curbing politically problematic speech, policymakers will be operating in a normative environment which provides fewer constraints on their ability to supply new changes to platform content standards, and that those policymakers are more likely to see supplying new rules as within their appropriate scope of policy intervention than their counterparts operating in a more 'laissez-faire' context.

2.4 The 'Power Plus' Argument

Having outlined the various components of my argument — the neutral, collaborative, and contested platform governance strategies, and the supply and demand factors potentially at play — the final move is to combine these two components and help explain variation across these three types. The thrust of the argument is summarized in the simple 'decision tree' seen in Figure 2.1.

The first necessary condition for changing the rules for platform governance involves adequate demand. If government rule-makers demand new rules — where demand can be affected by political or electoral motivations, and the preferences of constituents mobilized by civil society groups, and has not been tempered by interest group lobbying — than change is theoretically possible. If this demand does not exist, or if it is successfully thwarted by interest group lobbying or by interdependent influence or coercion from other states, than I expect for the 'neutral' strategy, where government does not significantly challenge platform private authority, to persist. If sufficient levels of demand are present, than I expect two supply conditions to increase the likelihood that a contested strategy will be pursued: sufficient power resources to supply the demanded change (understood as regulatory capacity minus any institutional constraints on that capacity) and the presence of an interventionist norm among rule-makers. If both of these factors exist, I expect that the state will contest platform private authority.

If only one of these two factors is present at sufficient levels during a regulatory episode, I argue that the likeliest outcome will be a collaborative arrangement. For instance: collaborative arrangements may occur if a state has the power to supply new rules, but rule-makers see those rules as outside of their appropriate scope of policy intervention due to a laissez-faire normative environment. Inversely, collaborative platform governance may also occur if rule-makers have the demand to intervene and see intervention as normatively appropriate, but do not have the power resources to foist binding rules upon companies, due to institutional constraints or a lack of regulatory capacity. In these kinds of instances, I expect governments demanding change to seek that change via collaborative platform governance. Finally, if prospective rule-makers have neither the power resources to negotiate changes through collaborative, non-binding agreements (even given their costs being lower than contested, binding ones, requiring comparatively fewer power resources) or an interventionist normative landscape, than I expect a neutral outcome, despite the prospective demand for change.

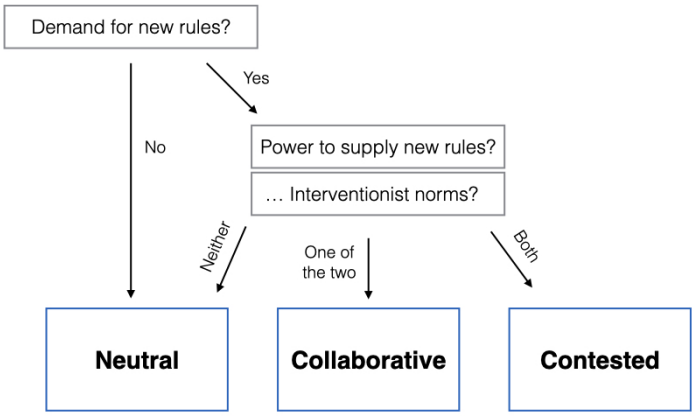

Figure 2.1: Decision tree depicting the core argument of the thesis.

2.4.1 Takeaways

This argument differs in a number of ways from the standard, power-driven Dreznerian story. In this context, where domestic rule-making is the main locus of contestation, power clearly matters, but not exactly in the way articulated by Drezner for the international arena. Firstly, domestic power resources are essential — but might be better understood as domestic regulatory capacity, with an important factor being the degree of control a government has over the legislative process. This introduces new political factors to consider for the growing and interdisciplinary platform governance literature: for example, does a government have an adequate grip on the legislative process (e.g. holding a strong parliamentary majority, and robust-enough coalition in democratic systems, or sufficient executive decision-making power) to implement new rules and meet demand for change? Or, in cases where a government faces transnational institutional constraints on its ability to make rules (for example, due to prior commitments made via trade agreements or international agreements, or broader institutional structures on rule-making within a regulatory harmonization zone like a regional single market), how do the

transnational power resources to overcome those constraints manifest themselves?

I call the argument a 'power plus' argument, because adequate domestic power resources are a necessary, but alone insufficient condition for a state to engage in contested platform governance. The binding commitments that underpin contested strategies of platform governance require not only the exertion of political capital and ability to withstand economic threats and coercion from firms, but can potentially pose reputational costs due to the normative ramifications of certain types of platform regulation. Different states are likely to perceive these costs differently, depending on the differing cultural contexts, and different positions in particular on the question of freedom of expression, and the appropriate amount of government intervention in the arena of information and communication. The implications of this argument are that some states may have the domestic/transnational power resources to contest platform authority, but due to normative and institutional factors, they may pursue a collaborative strategy instead. Collaboration may also be a strategy for states that would prefer binding rules for normative and institutional reasons (for example, a history of unsuccessful self-regulation in media or telecommunications sectors), but do not have the power resources required to pursue the contested approach.

The corollary of these arguments, and in contrast with Dreznerian, market-power-centric arguments in the platform literature (e.g. Srnicek, 2016) is that even small states with sufficient executive control over the rule-making apparatus may actually find it easier to pass new rules for platform moderation than large states marked by polarization and partisan gridlock. There is no clear reason why smaller states, if they have the adequate domestic power resources and regulatory capacity, could not pass domestic regulation that sets binding commitments for multinational corporations operating in their jurisdiction, if that change is demanded. As in other regulatory domains/industries, domestic regulation may be fiercely opposed by industry, but if a state has adequate demand and adequate means to supply changes to new rules, it should be able to — although if firms mobilize their larger home-state

governments to exercise diplomatic or economic coercion against that smaller state, that might become much more difficult for both demand and supply reasons.

2.5 Conclusion

In their overview of the emergence of non-state authority in world politics, Hall and Biersteker (2002, p. 6) discuss the work of the pioneering political economist Susan Strange, who argued that "non-state actors, such as enterprises, transnational social institutions, and non-governmental organizations, are increasingly acquiring power in the international political economy, and to the extent that their power is not challenged, they are implicitly legitimated as authoritative." This depiction fits fairly well within the understanding of tacit, tolerated, neutral platform governance described above. But interestingly, platform companies increasingly appear to be having their power challenged by state actors under certain conditions and in certain contexts. The goal of this chapter has been to extend a simple theoretical framework through which one can better understand this question, and to provide a stylized model for the dynamics of public-private regulatory contestation that occur given the decision by a state to shape platform policies and practices around user-generated content.

The argument presented is a modified supply and demand framework for regulatory change, informed by historical institutionalist insights as well as power-driven accounts in IPE. I argue that the interplay of state demand for new rules, the state's power to supply those rules, and the normative landscape within which government rule-makers operate in, all combine to lead to either neutral, collaborative, or contested strategies. The hope is that this relatively simple toolbox offers a vocabulary through which to analyze different modes of contestation around key regulatory episodes and case studies, and through which to help understand why — and how — certain internationally important policy initiatives succeed or fail. Additionally, it should help frame a better understanding of the historical evolution of platform regulation and governance in a cross-national perspective.

This argument, I believe, contributes a lens through which to start thinking through the emergence or non-emergence different forms of platform regulation for harmful content in a variety of contexts, including countries with differing degrees of democracy across the Global North and under-explored Global South. For example, using this model we can start identifying credible possible explanations for a number of regulatory developments. To provide an example which is not frequently discussed in the Eurocentric information policy literature: why have there been effectively no regional regulatory regimes for platform content emerging on the African continent, with the exception of in South Africa and Tanzania (where policymakers have used a limited broadcast media toolbox as a way to supply some new rules that relate to user-generated content on online platforms)? A detailed exploration could look at whether there may be insufficient demand from key interest-groups in certain countries (is the status quo, which largely consists of private regulation, seen as not great but good enough by policymakers, or has demand been depressed by lobbying from firms?), an inability to intervene due to regulatory capacity or domestic/transnational institutional constraints (e.g lock in effects due to trade agreements with Global North countries looking to protect their firms; pressure from platform company home states?), or perhaps other normative factors (such as various normative constraints limiting intervention)? Or to take another example, why has there been so much regulatory change in China, and how have policymakers been able to supply such rules which are stronger than anywhere else in the world? An exploration of this might discuss how China appears to feature very high demand for strong rules, given that leading political figures see content moderation as part of a broader information control strategy that is essential for the continued maintenance of Party control; the country also features significant regulatory capacity and resources, and policymakers see their efforts to supply those rules as highly appropriate, with few normative constraints that could tamper their intervention in the 'private sphere' of communication, even in the face of human rights harms. Furthermore, there are few institutional linkages to broader global structures in the regulatory domain of content moderation that might pose

costs for China to impose the types of rules it wishes, in contrast to other policy areas that are more strongly institutionalized at the global level.

Nevertheless, the argument presented in this chapter is still a relatively narrow one. Firstly, my framework is government-centric: I am primarily looking at contestation that occurs when government decides to intervene, and focusing less on firm strategies and civil society. The governance strategies discussed here are primarily around state-firm configurations. This is because the key point of interest is the development of government-led rules for platform governance, although civil society's role in influencing demand through its advocacy certainly plays a part of that. Other work has provided a far deeper assessment of the role played by civil society on related digital policy issues than can be provided here (Maréchal, 2015; Mueller, Kuerbis, and Pagé, 2007). Relatedly, the model looks less at the processes going on inside of companies, and the decision points at which they decide to make changes to their influential systems of private regulation on their own, separate (if such a process can ever be entirely separated) from the ongoing negotiation and contestation they face over those rules and practices from governance stakeholders across multiple jurisdictions. Given the current level of transparency that platform companies have about how exactly they make these policy changes (not much) and the way that they interact with researchers, it is not feasible to methodologically peer into the black box and get a better idea of state-firm interaction leading to specific changes. Partially for this reason, this model (and thesis) focus on the provision of public or public-private regulatory change.

Secondly, this chapter does not make much of an argument as to the specific types and character of the rules that are created within a contested or collaborative platform governance mode. Not all rules are equal, in the sense that some binding efforts may be far better crafted and far more effective at creating change than others pursued in other jurisdictions. Similarly, some contested platform governance strategies might yield rules that are more problematic and harmful in terms of human rights outcomes than others. These granularities will be unveiled through detailed case studies; the core of the matter that I am interested in exploring lies

more with the character of public and private authority and how it is asserted by governments and firms in the platform governance domain, and less with the specific provisions, structures, and mechanisms through which this occurs.

Finally, readers might be noting that this theoretical framework is centred predominantly on a single domestic or transnational policy arena, depending on the context (e.g. is the demand for, and supply of regulatory changes being pursued at the domestic level in a single country, within a group of countries like the G7 or G20, or regionally within the EU). But the framework says less to explain under which conditions a domestic set of rule changes yields a transnational impact — whether that occurs via firms taking the new rules and internalizing them across all markets, or via policy diffusion to other state jurisdictions. While this is an important angle I am interested in exploring, for the sake of this thesis I have opted for parsimony and a narrower scope to increase the feasibility of this doctoral project. I hope in future work to be able to further build upon this groundwork to look at these broader questions.

3

The Emergence of Platform Regulation, 1995-2020

Contents

3.1 Introduction

Any effort to understand the evolution of regulative institutions as they pertain to harmful content on major user-generated content platforms is necessarily limited by the fact that there is currently no reliable source to which one could look to observe these developments at a macro-scale. Unlike the existing efforts to

track and code all of the world's data protection (Greenleaf, 2019) and copyright (Herman, 2013) regulatory frameworks, there have been no comparable efforts to track what this thesis calls platform regulation. If one is interested in exploring various regulatory episodes involving public and public-private contestation over content moderation standards — and the various laws, regulations, and public-private initiatives that exist so far — it is unclear what the total universe of potential cases of interest looks like. From a macro-perspective, we still do not exactly know how many relevant formal and informal policy instruments are out there, and, longitudinally, how the global landscape has evolved since the founding of major platform businesses in the early 2000s.

This chapter represents a first empirical effort to provide a mapping of this formal and informal policy environment shaping online content regulation, a specific subset of the broader landscape of internet-related regulation. It presents an overview of what I define as platform regulation institutions — the formal and informal frameworks which set out rules for how platform companies conduct content moderation – across 'harmful content' issue areas over the past two decades. By synthesizing a number of existing resources and datasets, and building upon them with a combination of literature and comparative policy analysis, I attempt to provide an as-comprehensive-as-possible outline of the laws, legislation, co-regulatory institutions, and other standards-setting bodies relevant to this thesis. While this is a necessarily first effort, one limited by the paucity of existing datasets, I seek to combine this basic empirical description with analytical tools from the global regulation literature, including recent efforts to catalogue and map informal transnational governance initiatives and informal regulatory authority (Green, 2014; Roger, 2020; Westerwinter, Abbott, and Biersteker, 2021).

Furthermore, the chapter uses a few approaches from the extant regulatory politics literature to provide a more systematic overview of this regulatory landscape than currently available. In the first part of the chapter, I present an analysis of major trends in formal intermediary liability regulation, building out a rudimentary historical analysis that draws upon the best available existing data (especially the

Stanford World Intermediary Liability Map). This section outlines a universe of formal regulation and demonstrates a general increase in intermediary liability regulation more generally, and in the past five years, an apparent increase in the kind of platform-related contestatation that is the subject of this thesis.

The second part of the chapter discusses informal platform regulation, providing new insights from a new dataset on public-private transnational platform governance initiatives, following the best practices and coding scheme recently outlined by Westerwinter (2021). Using those criteria for scoring the governance functions and institutional features of these informal regulatory initiatives, I demonstrate that these initatives remain fairly scare and fairly informal, without many of the institutional features that have been seen in other, more established areas of international private governance. Additionally, I use the Westerwinter (2021) approach to provide a more systematic typology of the different existing forms of informal private platform regulation. Here, we also see an increase in collaborative informal initatives in the last five years that is worthy of further focused exploration.

The chapter concludes with a brief discussion of the formal and informal regulatory universe outlined through this chapter, and an explication of the case selection process that informs the empirical chapters that follow.

3.1.1 Scope

My goal is to provide a sketch of *the global universe of regulatory initiatives directly shaping how platform companies govern harmful content.* There are a few definitions needed to bound this mapping effort conceptually. None of these key definitions (what exactly constitutes a regulatory initiative, which types of regulatory initiatives specifically affect firm content moderation, and what should be considered harmful content) are uncontested, and it is inevitable that others might quibble with these definitions and the criteria for inclusion outlined here. Nevertheless, any data collection effort requires difficult judgements and boundary work, and I hope that this effort will provide a first step towards more comprehensive comparative policy research in this area.

My scoping definition has three elements. Firstly, by specific regulatory initiatives, I mean governance efforts that set rules, prescriptions, or best practices with proscriptions as to how companies should set and enforce rules around user-generated harmful content. I wish to consider the widest possible spectrum of regulatory types, including traditional 'command-and-control' domestic laws and formalized international treaties (Black, 2001), co-regulation or cooperative regulation efforts that involve significant competency sharing (Finck, 2017; Marsden, 2011), and what Abbott and Snidal (2009) call 'regulatory standards-setting' (RSS), the increasingly prevalent informal and voluntary governance agreements structuring the behaviour of firms and other internationally relevant private actors. Broadly, these can be categorized as 'formal' or 'informal' types of regulatory arrangements, with formal regulation codified by a domestic legislature or international treaty, and informal regulation effectively consisting of everything else. I am interested in both formal and informal arrangements that originate in any country, as both potentially can be impactful in shaping the rules for online content moderation (Gorwa, 2019).

The second element of my scoping definition, 'directly shaping how platform companies govern content,' is a more difficult one to precisely pin down. A platform's policies and practices could theoretically be affected by a wide range of different types of policies from domains like telecommunications law, media and broadcast regulation, and other general forms of internet regulation, such as network neutrality or common carriage provisions (Brown and Marsden, 2015). Much of this law proscribes general rules for internet intermediaries or online service providers (the specific terminology used varies), which thus affect platforms as part of the broader internet-related economy. While the emerging space of what has been called "online content regulation" — which might be considered a subset of internet regulation and the internet governance field more broadly (Papaevangelou, 2021) — is a relatively complex and emerging space, I believe that a distinction can be made between what is often called 'general liability' regulation and 'platform content regulation.' The difference comes down to the specificity of the regulatory targets and its aims: general liability arrangements, such as the EU E-Commerce Directive

(2000) or the US Communications Decency Act (1996) have very broad scope and might address a multitude of issues pertaining to legal liability; platform content regulation specifically seeks to affect how platforms conduct content moderation, directly naming platform companies as their target or crafting the regulation's inclusion criteria so they specifically include online platforms to the exclusion of other types of entities. While a baseline law might have some platforms in scope, and thus affect them (for example, by providing the status quo legal environment upon which they have built their moderation processes), platform content moderation explicitly seeks to have an effect on how platforms design their private rule-making systems, and is crafted with a direct intended impact on private platform authority.

Finally, 'harmful content' has generally been understood in content regulation discussions as content which is not necessarily illegal (and thus rendered criminal in national legal frameworks), but rather a broader category of content considered offensive or objectionable by governance stakeholders (Kuczerawy, 2018). In recent policy discussions, such as the United Kingdom's 'Online Harms' White Paper, the concept of harmful content has been defined very widely, to include everything from online bullying and the use of social media by criminals to child abuse imagery and the promotion of violence and violent extremist groups (Nash, 2019). Some scholars and civil society groups have suggested further extending the concept of harmful content to include potential harms to public health (such as vaccine misinformation) or to democratic institutions (electoral disinformation, voter suppression; see Bowers and Zittrain, 2020). Additionally, content that violates one's data protection rights or privacy rights could additionally be arguably said to be considered 'harmful' under such a definition, and if one took the concept to include economic, as well as potential social or political harms, than content that infringed upon intellectual property or copyright could as well. For the purposes of this chapter, I wish to advance a narrower definition of 'politically and socially harmful content' — *content purported to be depicting or supporting terrorist and violent extremist groups, hate speech, child abuse imagery, and disinformation* — thus excluding regulatory initiatives relating to reputational and economic harms, such as defamation, libel, and individual

privacy, as well as intellectual property, counterfeiting, and copyright. While these are important policy arenas, they involve, in my opinion, different stakeholder groups and regulatory initiatives than the politically and socially harmful content context does. Logan (2019), who divides the intermediary liability policy landscape into three separate issue areas: copyright, privacy,[1] and harmful content. The former two policy areas have been explored in depth in existing scholarship, which has considered, for instance, the fight against counterfeit goods online and musical piracy led by Hollywood and big business (Haggart, 2014; Tusikov, 2016), and the case law around the implementation of the 'right to be forgotten' in Europe (D. Keller, 2018; Kuczerawy and Ausloos, 2015). The politically harmful content area is an emerging, important, and understudied part of this broader internet regulation conversation (Papaevangelou, 2021).

In summary, I am seeking to map the following: formal and informal regulatory initiatives (formalized domestic and international legal instruments, domestic co-regulatory efforts, and private regulatory-standards-setting efforts) across the broad issue area of politically harmful content. Additionally, given this thesis' interest in patterns of institutional change, I wish to understand which initiatives specifically have provisions that implicate how platforms conduct their content moderation — and thus separate the baseline internet law which may have preceded platforms or does not fully engage with their governance challenges from the new forms of legislation which have emerged specifically to tackle user-generated content platforms.

[1]With the introduction of the 'right to be forgotten' in the European Union, data protection and privacy entered the intermediary liability conversation, as platforms (e.g. search engines) were required in certain cases to implement procedures to handle requests for takedown or 'de-listing' of content that could infringe on the reasonable rights to privacy of EU citizens, making privacy in effect the third area of online content regulation. See Kuczerawy and Ausloos (2015) for a full analysis of this area of intermediary liability.

3.2 Mapping Contested Platform Governance

3.2.1 Navigating Existing Data

There is very little comprehensive, comparative regulatory work that has been undertaken on the broad issue of online content regulation (Gillespie et al., 2020), and comprehensive data on regulatory and institutional emergence in this domain does not currently exist. However, there are some repositories of intermediary liability laws and court decisions more generally; in particular, a resource compiled by researchers at Stanford Law School's Centre for Internet and Society, the World Intermediary Liability Map (WILMap), is the best publicly available source of data on intermediary liability policies and content-related internet regulation. The WILMap is a volunteer-driven project with more than a hundred listed contributors, seeking to comprehensively map all "law discussing obligations and liability of online intermediaries due to (infringing) activities undertaken by their users," covering "almost one hundred jurisdictions in Africa, Asia, the Caribbean, Europe, Latin America, North America and Oceania" (Frosio, 2017, p. 3). The WILMap's main strength is its coverage: the team of researchers and volunteers assembling the data harnessed country and regional expertise to include all relevant laws they could. It contains "case law, statues, and proposed laws" across the wide range of content related policy sub-areas, from copyright, trademark and intellectual property infringement, to hate speech, defamation, terrorism, and more (Frosio, 2017, p. 3).

The WILMap has two major limitations: firstly, it is limited to statutory law, and thus largely excludes informal forms of regulation. Secondly, it exists as a public-facing online resource, and not a structured dataset that can be easily downloaded and analysed. However, I was able to obtain a copy of the WILMap dataset from Stanford's Centre for Internet Society with the goal of transforming it into a more easily accessible and publicly available dataset cataloguing platform regulation. I received the entire WILMap data in a JSON format; it contains 1081 entries catalogued with a variety of metadata, including the date, the country, topical tags (e.g "Hate Speech", "Terrorism," "Copyright," "Defamation"), and

categories for the type of development it is (e.g. "Regulation", "Court Decision," "International Agreement").

As a primarily legal repository, it contains a large number of court decisions, opinions, and policy documents. While court decisions are certainly important in shaping the overall implementation and ramifications of a regulation, I am primarily interested here in generally depicting the development and proliferation of those regulations in the first place. Filtering the data for all entries marked as "Regulation," "Legislation," and "International Agreements" yields 351 entries. These then were filtered by topic: I removed the copyright and intellectual property entries, as well as those pertaining to data protection and privacy, online fraud, consumer protection, and defamation. This left 129 different laws which formed the core of my formal regulation dataset. I accessed the text of each regulatory initiative, triangulating it with some other existing data on digital regulation compiled by the United Nations Conference on Trade and Development (UNCATD), removing laws which appeared clearly to be out of scope (e.g. a few copyright and intellectual property laws were still in the data; a few very general entries, such as national constitutions were also included), as well as some baseline laws for media and broadcasting that were passed before 1990, before the creation of the World Wide Web. This left a final sample of 111 regulations.[2]

3.2.2 Macro Overview

Efforts to map the emergence of global regulation often advocate for the importance of historical and longitudinal data (Green, 2014) to allow one to see broader trends and perhaps identify institutional juncture points (Campbell, 2004). I coded each regulation with the publicly available date of adoption (date that the law went

[2]Given the time constraints of this thesis project, I have been unable to engage in such an ambitious global data-collection effort from the ground up, and am instead relying on the best publicly available data as well as the categorizations used by the Stanford team. I cannot claim that this data represents a complete universe of all relevant legislation; and it may be skewed towards certain countries given the regional and language expertise of the collaborators that assembled it. Nevertheless, this data has yet to be used in this kind of platform regulation mapping exercise, and I believe that these limitations are outweighed by the illustrative high-level takeaways that can be presented in this chapter.

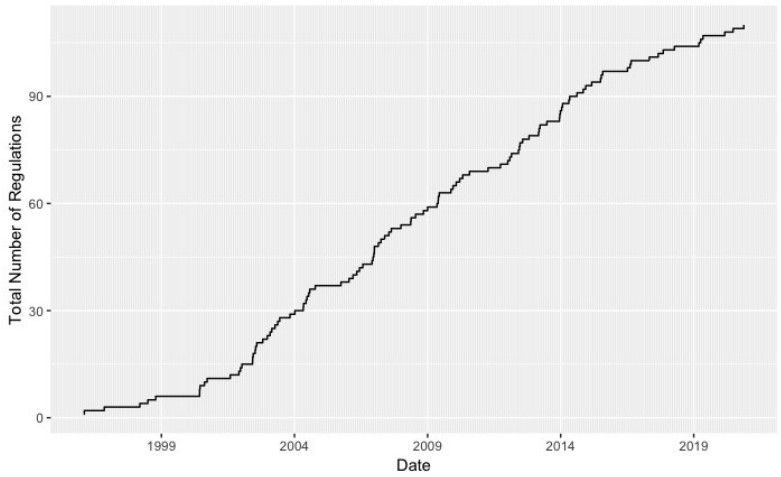

Figure 3.1: Longitudinal depiction of formal regulation in the Stanford WILMap data. Gridlines represent months.

into force; if not available, the date that the law was officially voted on or received official assent), which allows one to plot the total evolution of this sample over time. As seen in Figure 3.1, there has been a relatively steady growth in the number of global regulations implication harmful content online since 1990, with relatively flat growth in the first decade coupled with a sharper rise from about 2000/2001.

Before looking at this data in a more granular fashion, and zooming in on important regulations, countries, and specific time periods, a few more descriptive observations may be useful. Firstly, there is global distribution of these laws, with the dataset containing laws from 58 countries (mean: 1.86 regulations per country; median: 2). Most of Europe, the United States, and multiple Latin American countries are in the dataset, along with India, China, Taiwan, Malaysia, Singapore, the Philippines. Other included countries are Russia and Australia. Notably absent are most African nations, which generally do not have intermediary liability regulations pertaining to harmful content collected in the WILMap data — with the exception of Kenya, Rwanda, and South Africa. A fuller breakdown of the top countries is available in Table 3.1.

Table 3.1: Breakdown of top countries in WILMap data, by number of relevant formal legal frameworks. EU regulation is excluded here.

Country	Number of Laws
Russia	8
Turkey	5
Lithuania	4
Finland	4
China	4
South Africa	3
Rwanda	3
Kenya	3
India	3
Albania	3
United States	2
United Kingdom	2
Tanzania	2
Sweden	2
South Korea	2
Serbia	2
Philippines	2
Paraguay	2
Malaysia	2
Luxembourg	2
Ireland	2
Iran	2
Germany	2
Georgia	2
France	2
Czech Republic	2
Bulgaria	2
Brazil	2
Azerbaijan	2
Austria	2

In the following narrative overview, I describe the major time periods, trends, and type of legislation.

3.2.3 The Evolution of Formal Regulation

The Foundations of Intermediary Liability

Histories of the origins of intermediary liability law for the Internet often begin in the United States in the early 1990s, as American policymakers started to see

the challenges posed by various technologies that built upon internet protocols to provide various forms of multiparty communication (J. L. Goldsmith and Wu, 2006; Kosseff, 2016). These services, such as bulletin boards, e-mail, and personal web-pages, enabled not only direct peer-to-peer interaction, but also allowed for mediated mass communication, providing an opportunity for speakers to spread information at a scale that was once reserved only to those with access to established (and usually comprehensively regulated) traditional media gatekeepers. The first intermediary liability framework emerged in the mid-90s United States with the effort to remedy a paradox: online bulletin boards and hosting services like CompuServe, Prodigy, and America Online, which were becoming increasingly popular, were taking measures to set rules for what their users could acceptably say and removing pornographic material in an effort to provide 'family friendly' or otherwise curated services (Kosseff, 2019), in effect serving as the precursors of modern social platforms like Facebook that seek to provide a community-friendly experience for their users and their advertisers (Gillespie, 2018). However, due to the quirks of an American legal tradition grounded in the First Amendment, by conducting this early type of community moderation, bulletin boards were exposing themselves to potential legal risks, as courts had begun to interpret their moderation as actions that took on curatorial responsibilities as publishers or distributors with potential legal responsibility for the content posted by their customers (Kosseff, 2019; Wagner, 2016).

After a number of early cases where bulletin board operators were sued for defamation or other torts, two congressmen inserted a short clause into a piece of telecommunications reform regulation with the goal of creating an legal environment where companies would not be afraid to crack down on illegal or unsavoury content (Kosseff, 2019). Their short amendment, the Internet Freedom and Family Empowerment Act, was so ahead of its time that, according to the detailed history provided by Kosseff (2019), it was met with effectively no lobbyist or outside political influence, and in fact was almost totally ignored by commentators and the popular media following its adoption into law. It eventually would be codified

a Section 230 of the Communications Decency Act of 1996, and CDA 230, as it has commonly become known, provided operators of internet services with a 'safe harbour' through which they could conduct moderation without taking on legal liability (Chander, 2016). While this policy was supposed to initially provide a shield for family friendly, responsible moderation, it in subsequent years came to be interpreted extremely widely by the US courts, going far beyond what the framers anticipated to provide a legal shield for basically all types of online intermediaries against third-party lawsuits (Citron and Wittes, 2017; Kosseff, 2016).

CDA 230, as passed in 1996, became very influential as part of the European approach approach to regulating online services as well. In 1997, the European Commission published a communication on European commerce, which kicked off a regulatory process that eventually resulted in the E-Commerce Directive (2000/31, the ECD; see Julià-Barceló and Koelman, 2000). The ECD sought to harmonize European rules for 'information society services' provided by a wide range of different online intermediaries, from network operators (e.g telecommunications companies), search engines, web hosting providers, and social networks, with the goal to reduce diverging national standards for marketing, contracts, and the liability of intermediaries (Baistrocchi, 2002). Additionally, the ECD was meant to help remedy the question of jurisdictional conflict within the EU, where businesses may have faced legal uncertainty about which national rules should apply for cross-border services (Hellner, 2004).

The US model of intermediary liability follows a so-called 'vertical approach,' where different domains are covered via different mechanisms, Most notably, the procedures for online copyright are covered not by the Communications Decency Act, but by the more stringent Digital Millennium Copyright Act (DMCA) of 1998, which allows for liability for intermediaries that do not properly implement a system for managing complaints from copyright holders (Haggart, 2014; Meyer, 2017). In Europe, the E-Commerce Directive instead opted for a 'horizontal approach,' where all types of liability across different content areas are covered by the same framework. One important distinction between the EU and US approach is that

the ECD distinguishes between intermediaries that are 'mere conduits' and thus should have less responsibilities, and those that are more active 'hosts' and thus have more responsibilities (Kuczerawy, 2015). Article 14 of the ECD establishes a safe harbour for intermediaries that host user-generated content from third-parties as long as they (a) do not have knowledge of the illegality of content and (b) act to remove or restrict access to content once they obtain knowledge of that content's illegality (Angelopoulos and Smet, 2016). Article 14 thus established the conditions for what is commonly called a 'notice-and-action' scheme, with a high bar for intermediaries to be found criminally or civilly liable for the content of third-party users using their services: those intermediaries must receive notice of content they are hosting/transmitting from some other third party, and fail to act expeditiously on that notice (Kuczerawy, 2015). They receive so-called 'safe harbour' as long as they can show that they act upon notices in a reasonable manner.

The EU has two main mechanisms through which it enacts law: directives, and regulations, each of which allow for different types of regulatory harmonization in the European Single Market. While regulations are immediately applicable for all member states, directives *direct* member states to implement a law into their national legal frameworks within a certain period of time. The ECD is a directive, and thus was introduced, into the law of most member states, leading to 22 variations of the ECD which are captured in the Stanford dataset[3] (see Figure 3.2, which separates national implementations of the ECD from all other regulations).

The E-Commerce directive was part of an increasing movement towards many countries formalizing baseline e-commerce and intermediary liability frameworks, with other baseline frameworks adopted in June 2000 including also India's Information Technology Act and the Philippines' Electronic Commerce Act. Countries like Japan (2001), Brazil (2002), and Tanzania (2003) followed suit quickly; and in the EU, Luxembourg was the first member state to adopt the ECD, just two months after

[3]Some national laws are almost identical translations of the ECD (e.g. Italy); other countries, such as Romania, made changes that would still comply with the overall framework.

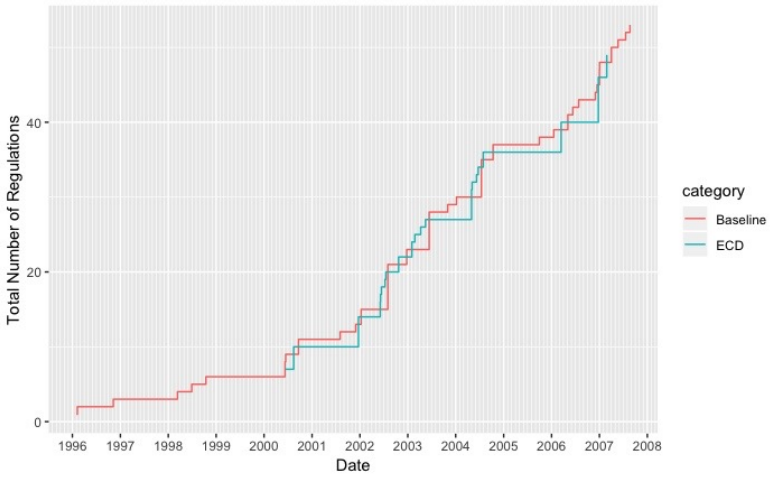

Figure 3.2: Longitudinal depiction of formal regulation in the Stanford WILMap data. Gridlines represent months; the blue line represents EU member state implementations of the E-Commerce Directive.

the Directive went into effect; the UK and nine other member states would do so in 2002, leading to a rapid proliferation of baseline legal systems for notice-and-action.[4]

Fighting For Control (or Freedom)?

The creation of systems of online liability did not resolve the fundamental question of how individual government rules and preferences would be reconciled over transnationally available material (Kohl, 2007); if anything, these systems heightened the debate over who exerted influence over online 'gatekeeper' intermediaries, and how this influence would be, and should be, exercised. While the battle over specific content rules practiced by companies — for example, whether online marketplaces in Europe, or anywhere for that matter, should allow customers to purchase Nazi memorabilia — were playing out largely through the courts as advocacy groups brought forth complaints (J. L. Goldsmith and Wu, 2006), less democratic countries were also seeking to use the intermediary liability toolkit to exert pressure and control over how their citizens could consume information online (Deibert et al., 2008, 2010).

As countries like Russia (2006) and Turkey (2007) followed in China's footsteps and established legal frameworks looking to 'regulate publications on the internet,'

[4]Adoption of the ECD would continue past the formal deadline of January 2002 for implementation set by the EU Commission. Germany passed its Telemedia Act, which implemented the ECD rules, only in 2007, and Belgium took until 2013. See European Commission (2002).

a lively academic and policy debate was kicked off about the political implications of internet-enabled technologies. US President Bill Clinton infamously stated in 2000 that trying to regulate speech on the internet would be futile, "sort of like trying to nail Jello to the wall" (Allen-Ebrahimian, 2016, n.p.), and yet China and Turkey both would successfully set up a licensing regime, where corporations wishing to operate online services that country would not just need to receive a license to do so, but would also need to follow certain moderation rules and procedures. Intermediaries that did not comply with removal requests for specific pieces of content (such as posts, news articles) made under laws like Iran's 2009 Computer Crimes Law or Turkey's 2013 Omnibus Bill No. 524 faced significant criminal sanctions and the possibility of having their services blocked by state-influenced or state-controlled ISPs (Stanford World Intermediary Liability Map, 2018). Even in India, the world's largest democracy, a formalized procedure for compelling the removal of specific instances of online content was adopted in 2009 (Stanford World Intermediary Liability Map, 2017).

This emerging process of what MacKinnon (2013) has called "networked authoritarianism" clashed directly against the US State Department's policy of "internet freedom," where the US government, driven by a host of economic and geopolitical considerations, sought to limit the global diffusion of strict internet liability rules designed to potentially curb the influence of grassroots democratizing social movements online (Powers and Jablonski, 2015). Nevertheless, an analysis of the Stanford data shows that it was not only less democratic countries that sought to shape the way that online intermediaries govern content. Issue based legislation has been common in high-income countries across a variety of policy areas. For example, as early as 1998, Ireland passed child safety regulation that implicated online services;[5] in 2006, Finland passed regulatory measures designed to 'prevent the propagation of child pornography.'[6] Brazil, Colombia, Paraguay,

[5]See the Irish 'Child Trafficking and Pornography Act, 1998', Stanford World Intermediary Liability Map (2017)

[6]See the Finnish Act 1068/2006, on the Measures Preventing the Propagation of Child pornography, December 2006, Stanford World Intermediary Liability Map (2017)

South Africa, South Korea, Russia, the Philippines, and the United Kingdom all in the past 3 decades passed legislation to combat child exploitation that had an 'online' element, criminalizing the online distribution of child abuse imagery. These specific issue-based intermediary liability laws generally placed some sort of requirements on online service providers, such as requiring them to report child abuse imagery that they uncovered on their networks, or in some cases, incentivizing them to deploy technological systems to filter and block such content (McLelland and Yoo, 2007; Powell, Hills, and Nash, 2010).

From Liability to Responsibility: Increasing Contestation

"We have arrived at the end of the beginning, the end of the regulatory beginning of the Internet," wrote Kohl (2012, p. 185) in 2012, an observation based upon the growing amount of regulatory complexity and variety that was starting to develop in online content regulation. While exact juncture points are difficult to pinpoint, it is evident that since around 2012, there has been a gradual shift in the liability landscape that eventually culminated in the rise of platform regulation. Frosio (2018, p. 1) has written that this shift represents a significant conceptual shift from "intermediary liability to intermediary responsibility." This development was probably the most acute in Europe, where policymakers, while working within the baseline constraints of the E-Commerce Directive, have for the past decade sought to layer on various issue-based voluntary responsibilities promulgated via various co-regulatory or self-regulatory initiatives (Frosio, 2018; Marsden, 2011). A fuller discussion of informal regulatory initiatives will follow in the second part of this chapter, but the central element to note for now is that online platforms — which by the mid 2010s had become established as large, profitable, and monopolistic transnational gatekeepers — were across a variety of content domains, across multiple jurisdictions, increasingly expected to fulfill a number of additional commitments to governance stakeholders (Gorwa, 2019).

By 2016, as events like the Brexit referendum and the election of Donald Trump to the US Presidency raised the public profile of content governance issues, specific

regulations began to be developed that sought more directly to influence and set guidelines for the content moderation policy practices of these companies, going beyond previous blacklist or takedown request based approaches. While many previous intermediary liability frameworks were horizontal, applying broadly to all forms of online content, and a wide range of different types of intermediaries (social media platforms, cloud providers, peer-to-peer messaging services, infrastructure providers, internet service providers), new laws were developed in some jurisdictions that specifically targeted platforms and how they handled content in specific political issue areas.

Contested Platform Governance, 2017-2021

As a host of new laws that specifically named online platforms as the regulatory target, and sought to directly affect how they conducted their moderation practices, commentators and scholars increasingly began referring to these policies as 'platform regulation' (Gorwa, 2019). In the previous chapter, I outlined a simple definition of 'contested platform governance' encompassing binding rules affecting specifically how platforms conduct content moderation relating to harmful content.

The German Network Enforcement Act (2017) set an important precedent as the first law to require mandatory transparency reports into the content moderation procedures of platform companies, and to mandate the creation of a separate complaints architecture through which user flags could be processed, was designed to only apply to very large social platforms above a threshold of 3 million users in Germany. Another piece of regulation in the Stanford dataset, the Singaporean Protection from Online Falsehoods and Manipulation Act (POFMA), specifically names its regulatory targets (YouTube, Facebook, Instagram, Baidu, WeChat, Twitter),[7] and attempts to force platforms to impose their content moderation processes to a form of government-led fact-checking and 'correction notices.'

The Stanford dataset contains 6 regulations that, upon close reading of their legislation, can be considered within the broader intermediary liability context as

[7]See the POFMA implementation guidelines, Singapore Government (2019).

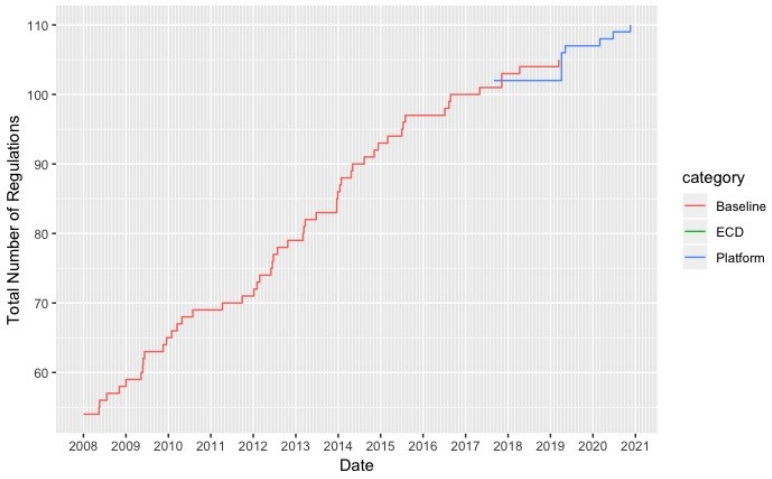

Figure 3.3: Longitudinal depiction of formal regulation in the Stanford WILMap data, 2008-2021. Gridlines represent months. The blue line depicts 'platform' regulation.

instances of contested platform governance, directly seeking to affect how platform companies moderate user-generated content in their jurisdiction: the German Network Enforcement Act (NetzDG, 2017), the Australian Abhorrent Violent Materials Act (2019), the Singaporean POFMA (2019), the Chinese government's Provisions for Governing Online Information Content (2020), the French Law Against Online Hate (Loi Avia, 2020), and the Austrian Communication Platforms Act (2020).[8] As seen in Figure 3.3, platform-focused efforts have emerged as a central subset of formal intermediary liability regulation since 2019.

The WILMap data also contains a few proposals which would be considered to be platform regulation under this definition but are not yet in force: new rules proposed in early 2021 by the Indian government (Intermediary Guidelines and Digital Media Ethics Rules), and in 2020 by the Pakistani government (Citizens Protection Against Online Harm Rules). If we were to add the other well-known draft regulations that have been discussed in the academic and policy literature in the past few years, the list would additionally include the UK Online Harms legislation, which has been in the works since an initial 'green paper' published

[8]One edge-case excluded here is the Malaysian Anti-Fake News law of 2018. The short-lived Malaysian law had the central aim of increasing individual liability for those who created or disseminated 'fake news'; it did not mention platforms, or specifically seek to affect the content moderation procedures of firms. For this reason, I do not consider it to be platform regulation following my definition. See ARTICLE 19 (2018)

in fall 2017 (Theil, 2019); the EU Terrorist Content Regulation, which has been actively negotiated since 2018 but has yet to go into force; and the Brazilian 'Fake News Law' (PLS 2630/2020) which was introduced in 2020 but has yet to pass through the full legislature. While a full discussion of the granularities of each of these formal regulatory arrangements is out of scope for this chapter, this set of initiatives provides the total universe of 11 possible cases of contested platform governance for this thesis to explore in depth.

3.3 Collaborative Platform Governance

The Stanford Data provides a helpful look at some of the macro-level trends in platform content regulation and can help guide a very brief intellectual history of major developments. Formalized, state-led regulation is no longer the only important form of regulation, however. As ample literature from regulatory politics and global governance scholars has documented (Abbott, Green, and Keohane, 2016; Büthe, 2010), a wide variety of private and informal governance has become increasingly common in both the domestic and international arenas, with governments increasingly relying on informal arrangements to achieve certain aims (Roger, 2020), and non-governmental actors deploying various private standards setting arrangements for economic (e.g. signalling their credible commitments to policymakers and to publics)[9] or for normative reasons (Renckens, 2020; Tusikov, 2017). The complete regulatory landscape that needs to be addressed thus includes not only traditional forms of regulation, but a host of voluntary arrangements, public-private partnerships, industry-specific measures, and other collaborative arrangements with varying distributions of governance roles (Abbott and Snidal, 2009; Mattli and Woods, 2009; Zürn, 2018).

The Stanford Intermediary Liability dataset unfortunately only has good coverage of formally codified laws and regulations, and does not include organized

[9] As Abbott, Green, and Keohane (2016, p. 248) note, "voluntary standards rely on incentives such as consumer demand, reputational benefits, avoidance of mandatory regulation, and reduced transactions costs" but the varying types of actors involved in private regulatory efforts — ranging from civil society groups to firm associations — can be motivated by varying blends of normative and interest based factors.

co-regulatory initiatives (such as codes of conduct developed by a mixture of government and firm actors) or private regulatory organizations of varying forms, which are increasingly being recognized as an important element — and indeed, perhaps the central developing trend — in online content governance (Douek, 2020; Gorwa, 2019). For this reason, a full mapping exercise like that undertaken in this chapter needs to also provide a holistic picture of the private and informal governance initiatives that influence actor behaviour in the policy domain in question.

3.3.1 Mapping Private Regulatory Organizations

Global governance researchers have consistently created datasets of private regulatory organizations in order to glean some empirical insights into institutional development over time, whether the topic of interest be the development of these organizations in general or pertaining to a specific issue area, such as international climate politics (see Westerwinter, 2021, for a full overview). The largest existing dataset of private regulatory organizations is the 'Transnational Public-Private Governance Initiatives in World Politics' (TGIWP Westerwinter, 2021), contains more than 600 governance initiatives across a variety of topics. However, it has only a handful that relate to technology policy matters, and upon examination, does not have any coverage of platform governance related initiatives. Regardless, it provides an overview of the best existing practice for creating datasets, and a coding framework which can be followed when adding to this data.

Trying to empirically map all occurrences of informal regulation globally would be a empirically and conceptually difficult task, so Westerwinter (2021) takes a slightly narrower approach by limiting the scope to 'public-private' informal regulatory initiatives that feature a blend of actor groups (so not just internal firm self-regulation) and at least a minimum degree of institutionalization (so that public information about the initiative exists). The data seeks to map out global 'transnational governance institutions,' which are defined as "institutions that 1) involve at least one state and/or IGO, one business actor, and one civil society actor; 2) are transnational in terms of their participants and scope of activities;

3) perform tasks that are related to governing transnational problems; and 4) are institutionalized to the extent that they provide a basis for regular interactions among their participants" (Westerwinter, 2021, p. 141).

This definition follows the direction taken by groups of leading global governance scholars who have created datasets looking at varying aspects of private governance in areas like global climate policy. In an overview of datasets created by scholars including Abbott, Green, and Keohane, and Fransen, Shalk, and Auld, Renckens (2020, p. 664) identifies dataset creation efforts generally having three components:

> First, they identify the schemes as organizations, which implies that loose networks or informal schemes are not included. Second, they select schemes that are private, in the sense that they are dominated or controlled by private actors. This means that public voluntary schemes, such as the United Nations Global Compact, are excluded. Third, they focus on schemes with a regulatory focus, thereby ruling out schemes whose functions include non-regulatory activities, such as information sharing or capacity building.

Westerwinter's definition is broader, in that it includes any organization that features a mix of actors (even if they are steered or controlled by state actors, with private actor involvement) and that it includes all schemes that are "related to governing" (Westerwinter, 2021, p. 141), including capacity building, information sharing, and other non-explicitly-rule-setting activities. This definition is attractive as it is often difficult to discern where exactly the governance boundaries of certain platform regulation-related private initiatives are (what exactly should be considered explicit rule-setting is more ambiguous here than for example the explicit creation of product standards, and private organizations appear to be frequently secretive about exactly what their activities are), and also because many of the public-private transnational regulatory initiatives appear to have more state involvement than might be expected in other policy domains.

Creating a Public-Private Platform Regulation Dataset

Taking Westerwinter (2021) as the current best practice for data collection and coding in this area, I sought to follow the same methodological approach to data

collection, coding, and analysis that is outlined in their supplemental materials and codebook. Firstly, I looked to unearth as many private regulatory initiatives that fit the above definition as possible, doing this through by looking at the leading literature on private regulation by platforms, and building off a prior mapping effort for the EU done in Gorwa (2019). This was supplemented by additional desk research that looked at initiatives mentioned in primary policy documents and grey literature from civil society groups and digital rights organizations. The primary mechanism to collect industry initiatives was via the public 'newsroom' pages through which they post product and policy announcements; I sought to triangulate government-steered initiatives which were not publicly acknowledged by companies via scholarly and policy literature and newspaper keyword searches. However, it is possible that government-led initiatives (especially domestic ones, with the participation of multinational companies that made them weakly transnational) that were only publicized in the host country (and in the language of the host country; I was only able to locate documents via searches in English, French, German, and Polish) were overlooked here.

The result of this data collection, which occurred in early 2021, unearthed 26 possibly relevant initiatives. These initiatives were then coded along the standard scheme established in the Westerwinter (2021) dataset. Along with the basic descriptive characteristics (name, year of inception), all publicly available material was parsed to code each initiative across three broad categories.

The first question set out in the Westerwinter approach is to determine what kind of actors were involved in the initiative: firm, civil society, or international organizations. To code this variable of actor participation, I looked to see if the publicly available documentation describing an initiative (its website, if any, as well as press releases, annual reports and more) listed the participation of a specific state or sub-state unit (e.g a regulatory authority, government department); civil society broadly construed, including non-governmental organizations, transnational advocacy networks, or research institutions; or platform companies or their current employees. These were coded with a binary 0-1 variable for each category, meaning

that the presence of just one firm in an initiative featuring, for example, many civil society organizations, would lead it to be coded with a 1 in both the civil society and firm actor presence categories. Importantly, for this exercise I coded only the latest available information pertaining to the organization in its latest form, and did not seek to track longitudinal development within organizations/initiatives over time. This means that an initiative like the Christchurch Call to eliminate terrorism and violent extremism online, which began with a mix of state actors (New Zealand and France steering, multiple other state signatories) and industry, but in September 2019 added an advisory body of civil society groups, was coded in early 2021 as containing input from all three actor groups. At this step, initiatives that only ticked one actor box (such as the Manila Principles on Intermediary Liability, which only had publicly listed engagement from civil society groups, or the Interim Codes of Conduct developed by the UK government in the lead up to new Online Harms legislation, which do not list any active participants or adopters in industry) were excluded from the next steps of the analysis as they no longer met the multi-stakeholder criteria for inclusion outlined by Westerwinter.

Secondly, I sought to understand what kind of governance functions the initiative purported to fulfill, based on the seven categories outlined in the TGIWP data: agenda-setting, standard-setting/rule-making, standard/rule implementation, monitoring, funding, capacity-building, and knowledge creation/information sharing. Assessing all the publicly available documentation I could find, I following the coding guidelines laid out by Westerwinter (2021), I sought to provide a best estimate of whether each transnational governance initiative had a stated: *agenda-setting goal* (understood as "campaigning to increase the awareness of specific target actors for a given issue or problem"); whether it *created standards* ("These can be technical standards, codes of conduct, guidelines, or any other form of standard that is geared toward changing the behavior of some target audience" varying in their formality and level of obligation); whether the initiative was involved in the *implementation* "of existing rules and standards" by being "involved in the design and implementation of activities and programs that are geared toward the implementation of and ensuring

the compliance with some international rule or standard" (all quotes from p. 12 of the online appendix in Westerwinter, 2021); and whether the initiative was involved in *monitoring* whether some target group (possibly the members of an initiative, but also possibly other actors) adhered to some kind of rule or norm. Additionally, I sought to assess whether the initiative was engaged in *funding* specific outcomes, such as specific projects or research as part of its activities (broadly construed to include "one-off fundraising campaigns, sponsorship, donations, or permanent institutional support" for either "material outputs, such as roads, wells or schools, or immaterial outputs, such as knowledge or research," whether it engaged in *capacity building*, efforts to build the capabilities of any group of actors; whether it sought to *create knowledge,* to develop "new thinking, research, expertise, ideas, and policies that are related to or concerned with transnational problems" by facilitating "exchange, sharing, and dissemination of information and knowledge and networking among its participants as well as participants and actors that are not participating in the TGI," and finally, whether it engaged in *service provision*, delivering a specific good or providing a specific service, such as distribution of a resource (all quotes from p. 13 of the online appendix in Westerwinter, 2021).[10]

Finally, I coded each initiative to provide an overview of its institutional design, based upon the eight-point categorization in the TGIWP data. I looked at all available public information for the following: a founding document, specific behavioral obligations for their participants, a secretariat (and whether this secretariat was independent, or provided by another organization or actor), a forum for regular meetings of participants, defined monitoring mechanisms to uphold membership, defined enforcement procedures against members not meeting the criteria of membership, a dispute settlement mechanism to adjudicate between members, and specific procedures for making decisions (such as a board or governing body with clearly delineated voting rules).

[10] "The service may be provided to the members of the TGI or to a larger group. Especially in the area of development and health, many TGIs are initiated to provide services, such as the distribution of drugs or the construction of wells or schools" (Westerwinter, 2021, online appendix, p. 14).

Given resource limitations (a lack of additional coders to perform inter-coder reliability checks), this was a rough process, and there were some edge-cases between categories in some instances. Nevertheless, I believe that using this coding scheme provides a helpful snapshot of the informal regulatory landscape for online content regulation, and in future work, I hope to expand upon this work in a larger scale and more robust fashion. A fuller exposition and description of the coding process and its limitations is available in Methods Appendix A.

3.3.2 Macro Overview

The result of this coding process was a small dataset of 23 public-private platform-content related transnational governance initatives (TGIs), or possible instances of collaborative platform governance. The earliest initiative in the dataset was founded in late 2008: the UK Government's Council for Child Internet Safety, which brought together government ministers and staff, representatives from telecommunications companies and social media platforms, and American and European child safety organizations to negotiate child safety guidelines and design elements. About a month later, the Global Network Initiative (GNI) was created to develop best practices for how internet companies should respect user rights (with the key question of how to deal with governments seeking access to user data), with Yahoo, Google, and Microsoft as founding industry members, joined by a group of NGOs, investor groups, and academics (Maclay, 2010). The most recent initiative in the dataset was TikTok's 'European Safety Advisory Council,' founded in March 2021 with a small number of individuals from European civil society and a mix of TikTok's policy employees working on harmful content issues.

While the total sample is small,[11] there has been a notable uptick in the number of initiatives that have been created since 2015. As seen in Figure 3.4, after a few early initiatives were founded in 2008-early 2009, there was a 3 year period

[11]The sample is small, but in line with the number of transnational private governance schemes academics are studying in areas like climate change generally or sustainable agriculture more specifically (Renckens, 2020); these initiatives have additionally do not appear to be not part of the major global governance data collection efforts looking at private governance (Westerwinter, 2021)

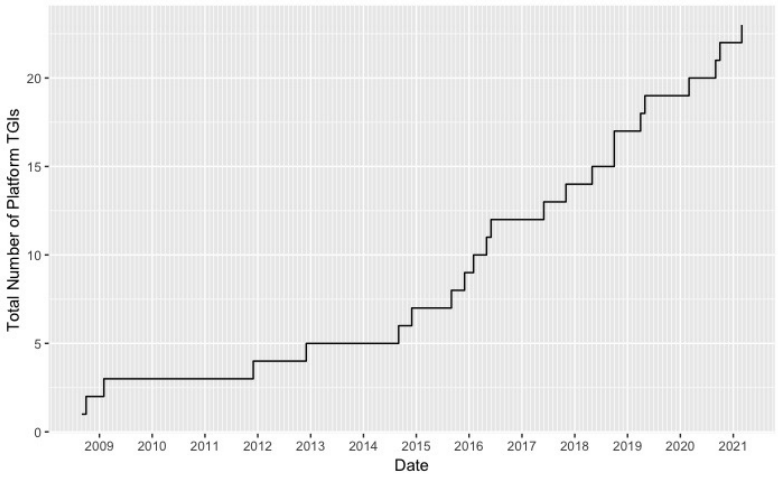

Figure 3.4: Longitudinal depiction of informal regulation for online content platforms in the newly created TGI data. Gridlines represent months.

with very little movement, with the 5th initiative being founded in late 2014. Since there, there has been a sizable uptick, with close to 20 new initiatives being created in the last 5 or so years.

Upon closer examination of the results from the coding effort, it becomes apparent that this policy space is not very formalized in terms of the institutional structures these private initiatives employ. Out of the 23 initiatives, slightly less than half have a founding document (see Figure 3.5). 10 initiatives create some sort of obligations for their members,[12] seven have some kind of recurring forum with decision-making power and a representative mix of participants, and five have a secretariat to support the initiative's efforts. (Of these five, only one, the Facebook Oversight Board, has a fully independent secretariat run by the initiative itself; the rest have a secretariat supplied by a government or industry actor). While specifying obligations explicitly is fairly common, far fewer initiatives actually specify monitoring procedures for ensuring that those obligations are met, and only one initiative (the Global Network

[12] According to the online appendix in Westerwinter (2021), "The variable Obligation is coded 1 if a TGI has one or several of the following characteristics: 1) the participants in the TGI are expected to behave in line with a specified set of rules, principles, recommendations, or guidelines created by the initiative; 2) the participants in the TGI have specified duties, tasks, or other obligations that they are expected to fulfil; 3) the participants in the TGI are expected to satisfy a specified standard or other criteria to become or remain participants of the initiative; and 4) the participants in the TGI are expected to execute or implement decisions of the initiative. If none of these four criteria is met by a TGI, Obligation is coded 0."

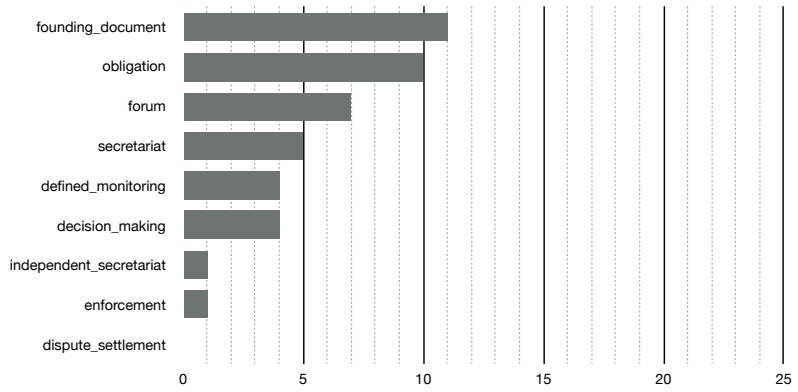

Figure 3.5: Breakdown of institutional elements in platform TGIs, following the Westerwinter coding scheme.

Initiative) specifies an institutional body that can enforce compliance and impose penalties to backsliders or non-implementers. Structured decision-making bodies (e.g. boards or executive committees) with publicly listed formal decision-making procedures (e.g. voting procedures) are additionally rare, with only 4 initiatives (the Global Network Initiative, the Facebook Oversight Board, the We Protect Global Alliance, and the Global Internet Forum to Counter Terrorism) specifying some sort of structured decision-making process. No initiative mentioned an institutional design for adjudicating or handling disputes between its member organizations.

Following the coding scheme for transnational regulatory initiatives in all aspects of global governance established by Westerwinter (2021) adds some additional insight into the specific governance functions that these initiatives fill in the platform regulation ecosystem. As Figure 3.6 shows, the most common contribution of these kinds of informal regulatory arrangements are for the most part 'softer' framing and agenda setting inputs: 3 out of the four most common governance functions pertain to capacity building (effectively all of the initiatives), knowledge creation and dissemination, and agenda-setting. 17 out of the 23 initiatives were involved in creating some sort of standards, codes of conduct, guidelines, or best practices, although these may vary in their significance and actual policy impact. Publicly defined functions for implementing and monitoring standards are comparatively much rarer in the collected initiatives, lending credence to

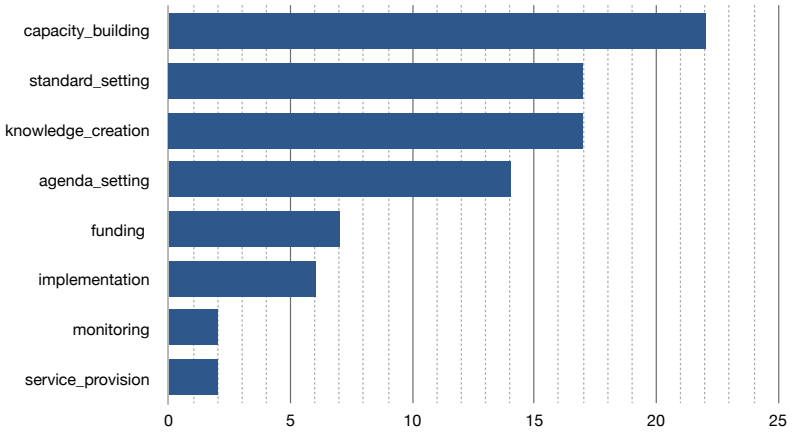

Figure 3.6: Purported governance functions filled by platform TGIs, following the Westerwinter coding scheme.

critiques of these efforts by journalists and digital rights groups that these platform-related organizations lack accountability mechanisms and 'teeth' in many cases (MacKinnon and Pakzad, 2018).

This more granular description of multi-stakeholder platform governance initiatives provides more detail that can build upon recent work on the topic (Douek, 2020; Gorwa, 2019), and allows one a preliminary evidence base upon which some descriptive conceptual arguments about the state of play can be made. Interestingly, this analysis, in combination with the formal overview presented in the first part of this chapter, demonstrates that informal means of transnational regulation are more plentiful — and have existed for longer — than direct state 'contested' efforts seeking binding rules for how platforms govern harmful content. The 'collaborative' mode appears to have been the popular dynamic for many policy issues at the transnational level and in the European Union, where the European Commission has long sought to apply a unique 'co-regulatory' strategy to digital services (Marsden, 2011).

How should the various multi-stakeholder initiatives seen here be understood and organized? In past work, I have simply described these initiatives based on their actor groupings (Gorwa, 2019), but the TGIWP coding framework provides an opportunity to make some more detailed observations. One simple heuristic that can be created using the Westerwinter (2021) framework is what one might call a governance score and an institutional score: adding up the amount of purported

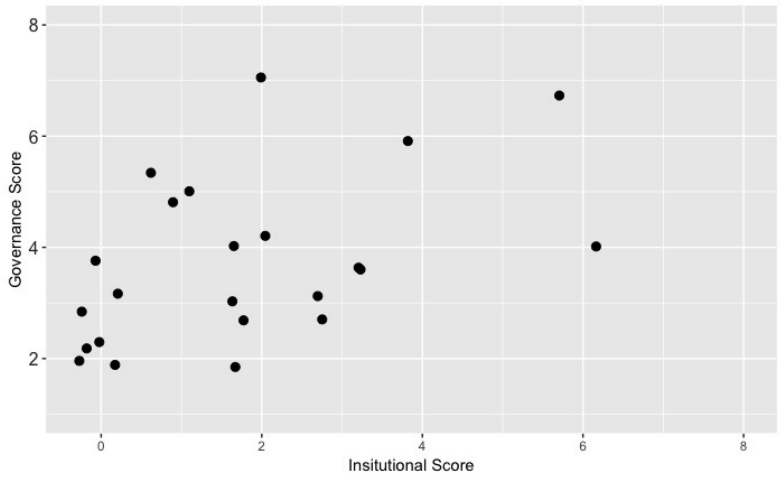

Figure 3.7: Jitter plot of the 'governance score' and 'institutional score' variables as coded. Each point represents an initiative, and is 'jittered' around the grid to allow one to visualize multiple observations at the same points.

governance tasks an initiative seeks to fill, and the amount of different institutional characteristics that it displays, to create a heuristic that represents both dimensions. This is a rough measure, given that the Westerwinter scheme seeks to map, at a macro-level, large-grained observations about many TGIs, and does not contain specific 'case studies' into each initiative. The coding scheme thus only measures if an initiative, for example, states that it creates some sort of standard, and not the actual impact of those standards. Similarly, an initiative might have a stated institutional framework for monitoring members, but that framework might be weak or (according to some) ineffectual. Nevertheless, looking at these scores provides a snapshot of this institutional landscape that is useful to make some additional observations.

Firstly, most initiatives fulfill relatively few governance functions and lack many institutional features. There are 6 initiatives which have an institutional score of 0 (see 0, on the X axis in Figure 3.7), and 8 that have either an institutional score of 1 or 2 (but may serve a correspondingly large number of governance functions, see the Y axis). This graph also highlights some outlier organizations that have a high number of institutional features and high number of stated governance functions (the GNI = 6, 7; the We Protect Global Alliance to Combat Child Abuse = 4, 6 in Figure 3.7), and a handful that might be said to have a 'governance gap', with a

high number of governance functions but comparatively few formal institutional checks (the Global Internet Forum to Counter Terrorism = 2, 7).

3.3.3 A Platform TGI Typology

Building upon this classification and coding effort, one can advance a typology of the different types of platform TGIs, based upon their groupings of actors, their governance functions, and their institutional characteristics (Table 3.2). The first category could be called 'Platform Advisory Boards.' These are TGIs that are created by platform companies with their own resources, and feature a selected group of civil society participants (either NGOs, or individuals representing a group of NGOs) that may be consulted by the platform company to provide specific feedback on certain policies or practices. They are generally extremely informal in structure (generally with an institution score of 0, given their lack of clear charter, voting procedures, or other mechanisms) and fill very few stated governance functions, other than perhaps seeking to increase the capacity and knowledge of certain industry actors on certain issues. The major exception to this category is the Facebook Oversight Board, which is also an advisory board of sorts, but one that has been also tasked with providing a specific service (adjudicating on content appeals referred to them by Facebook), and as a result, has also been created with a number of institutional elements (founding charter, independent secretariat, formally independent financial status, clearly delineated voting procedures) that are not present in the other initiatives.

Table 3.2: Informal Platform Governance Initiative Typology.

Type	Actors	General Characteristics	Examples
Platform Advisory Boards	Firms (steering) Civil Society (participants)	Governance & institution scores close to 0	Twitter Trust and Safety Council, TikTok Content Advisory Council;
Domestic Issue-Based Task Forces and Codes	Government (steering), Firms (participants)	Low institution score; low governance score but involved in standard setting	German Task Force Against Illegal Online Hate Speech, UK Council for Child Internet Safety, EU Code of Practice on Disinformation;
Multistakeholder Talking Shops	Government, Firm, Civil Society (steering varies)	Low institution score; low governance score, mainly involved in capacity building and knowledge transfer	EU Internet Forum, IGF Dynamic Coalition on Platform Responsibility
Transnational Capacity Builders	Government, Firm, Civil Society (steering varies)	High governance score, low-to-high institution score	Global Network Initiative, Global Internet Forum to Counter Terrorism, WePROTECT Global Alliance

The government-led counterpart to these advisory boards might be called "Task Forces." These are initiatives, often set up for a short period of time and around a politically pressing issue area (such as hate speech, disinformation, or child safety) by a branch of government (or a transnational actor like the European Commission) to build capacity and to negotiate voluntary rules together with industry. They are generally very informal institutionally, controlled by a single government actor which 'runs the show' (obviating the need for a charter, formal forum for discussion with voting procedures, etc), but can still have quite important policy ramifications, as the codes or guidelines for actor behaviour that they seek to promulgate in a

jurisdiction may lead platforms to make important changes to their policies locally or even globally — and all without a formal legislative or legal process, leading critics in civil society to occasionally dub such 'voluntary' regulatory initiatives with the moniker 'shadow regulation' (Fiedler, 2015; Frosio, 2018).

The third category, Talking Shops, are TGIs that bring together a host of different actors to discuss policy developments and to get certain issues onto the agenda. While some, such as the Dynamic Coalition for Platform Responsibility at the annual UN Internet Governance Forum (or conversations at the IGF more broadly) are relatively inconsequential, not creating tangible outputs, others, such as the EU Internet Forum, which has since 2015 been bringing together industry with security-focused officials from across the EU to discuss terrorist and violent extremist content in a secretive and closed-door set of discussions (Fiedler, 2015), are more impactful. These Talking Shops can provide an avenue through which states can levy informal pressure and seek to influence firms' policy agendas and perhaps develop new processes and systems (such as the Shared Industry Hash Database for terrorist content, which was created at the impetus of the EU Internet Forum and eventually scaled into the stand-alone Global Internet Forum to Counter Terrorism; see Gorwa, Binns, and Katzenbach, 2020). These types of TGIs do not create standards, but rather focus on knowledge sharing, funding, and agenda-setting; if they do seek to create standards, they may lead to formal initiatives or morph into code of conduct focused Task Forces.

Finally, the last category, which I call Transnational Capacity Builders, are the most formalized and (potentially) consequential of the bunch, and are the closest that the platform regulation domain has to the traditional understanding of public-private regulatory organizations in the global governance literature (Abbott and Snidal, 2009). These organizations pursue a broader set of governance functions (including the provision of services, like shared industry hash databases), the creation of new codes of conducts and guidelines, and monitoring and implementation structures. For example, the We Protect Global Alliance, an initiative battling child abuse content, lists 98 government members, 51 companies, and 52 civil society organizations that

have formally applied to be members, and its multi-stakeholder board helps shape the organization's advocacy, rule-setting, and policy work. They have experimented with a number of different voluntary initiatives well-established as strategies in the corporate governance and corporate social responsibility literatures (Fransen, 2012; Hofferberth, 2019), including the development of industry voluntary codes, best practices, and implementation guidelines, and the development of 'trustmarks' that members can display publicly if they fulfill certain criteria.

3.4 Conclusion

In this chapter, I sought to provide an overview of platform regulation as it has evolved as a sub-dimension of intermediary liability in the past three decades. This effort, like any data-driven analysis, has faced a number of important limitations. Firstly, the formal analysis provided in the first section of this chapter relies on the Stanford database, which may be the most comprehensive publicly available source for online content regulation, but still has a few notable shortcomings. Firstly, it has been assembled through country case studies, with legal experts in each country adding entries and providing short summaries of the landscape in each country. While this is helpful for getting a general snapshot of liability frameworks around the world, it may potentially overlook certain countries and regions, and alas likely errs towards providing an over-inclusive snapshot of all the policies that might have an impact on intermediary liability. These are important qualifiers to the descriptive inferences made about the 111 regulations analysed here; I cannot claim to be analysing the complete landscape, but rather, what I hope is an indicative sample that nevertheless provides a starting point for future work. The same is true of the informal regulation overview presented in the second half of the chapter: while I have done my best to obtain as many voluntary and informal regulatory initiatives, it is possible that my sample has a Global North, English-language bias, and that other relevant initiatives exist. One strategy that has recently been used to ensure maximal coverage in these kinds of regulatory mapping exercises is the use of surveys that are filled out by individual country or regional experts

(Euchner, 2020); this kind of approach could be used in the future to supplement this mapping effort of both informal and formal initiatives in the future to foster the creation of a public dataset for use by other researchers. Nevertheless, despite these limitations, the data presented here remains the best and most comprehensive look at these kinds of intermediary liability rules to date.

3.4.1 A Platform Regulation Universe

To provide an as-complete-as-possible universe of cases that this thesis can analyse as focused cases, we can consider the entirety of the formal regulatory initiatives coded as platform regulation in the first half of this chapter.

Table 3.3: Current universe of formal 'platform regulation' following my definition.

Country	Name	Date
Germany	The Network Enforcement Act (NetzDG)	01.09.2017
Australia	Sharing of Abhorrent Violent Material Act (AVM)	06.04.2019
Singapore	Protection from Online Falsehoods and Manipulation Act (POFMA)	08.05.2019
China	Provisions on the Governance of the Online Information Content Ecosystem	01.03.2020
France	Law Against Online Hatred (Loi Avia)	24.06.2020
Austria	The Communication Platforms Act (KoPl-G)	18.11.2020
EU	Terrorist Content Regulation (2018—)	Draft
UK	Online Harms Regulation (2017—)	Draft
Pakistan	Citizens Protection Against Online Harm Rules (2020)	Draft
Brazil	Freedom, Liability, and Transparency on the Internet Law (2020)	Draft
India	Intermediary Guidelines and Digital Media Ethics Rules (2021)	Draft

Additionally, the 23 instances of collaborative platform governance via an informal transnational governance initiative are cases of interest. The 10 initiatives which fulfill the greatest number of governance functions can be seen in the table below.

Table 3.4: Current universe of informal 'platform regulation,' in collected data on informal initiatives.

Name	Year Founded	Year Ended
Global Network Initiative	2008	
Global Internet Forum to Counter Terrorism	2017	
WePROTECT Global Alliance	2016	
CEO Coalition	2011	2015
Tech Against Terrorism	2017	
WePROTECT	2014	2016
Facebook Oversight Board	2020	
EU Hate Speech Code of Conduct	2016	
Global Alliance Against Child Sexual Abuse Online	2012	2016
Christchurch Call	2019	

3.4.2 Case Selection Considerations

Case studies are commonly used to drive analyses in political science and international relations. In the early 2000s, they become increasingly contentious as a method, with quantitatively-minded scholars raising concerns about the generalizability, replicability, and general value of small-N case studies, especially given the possibility of sampling bias (Gerring, 2004). However, on the flip side, case studies allow for far more in-depth analysis of variation at a closer level, allowing one to move beyond the abstraction and assumptions embedded in datasets. They can also permit one to more carefully examine causal mechanisms than correlation-driven large-N work (Campbell, 2004, p. 79). They remain a widely accepted strategy in institutional and regulatory analyses, as long as one is careful about not making over-broad claims based off of narrow cases and remains humble about their possible shortcomings.

Given this thesis' aim to better understand the conditions under which actors seek to obtain and supply changes to the transnational rules governing platform content moderation, ideal cases would be able to address both formal and informal regulation, and would have gone fully into effect so that as-full-as-possible spectrum of negotiation and implementation could be considered. Additionally, even though this is a rapidly developing area, given the time constraints facing this thesis, the

regulatory episodes leading to contestation between platform and state authority should have concluded by a time that provided the adequate number of months needed to conduct the necessary research. Only being able to select regulations that had gone into effect by the end of 2019 leaves a total of 3 formal regulations and 9 informal regulatory initiatives.

Given the number of chapters and cases that can be reasonably investigated in depth in a doctoral thesis, I cannot address all of these cases. While a random strategy would be potentially desirable if there was a larger total sample, given the relatively small number of potential cases, this study will require a purposive sampling strategy (Gerring, 2006). One commonly articulated approach is to look for either 'most likely' or 'least likely' cases to test one's causal mechanisms and arguments: picking a selection of cases that appear to most perfectly fit the theoretical explanation advanced in the thesis, and a selection of cases that do not. Another strategy frequently deployed in the regulatory politics literature is to deploy what has been called a 'diverse' set of cases (Gerring, 2006). In this approach, a range of cases that represents the full possible range of variation in the causal mechanism and the combination of variables that is in play. In this sense, the sampling strategy combines features of the 'most' and 'least' likely models of case selection (Kellerman, 2019), and highlights different components of the argument. For reasons of feasibility in this doctoral project, the cases need to be researchable in a doctoral time-frame, and would ideally provide some kind of within-case variation (between collaborative/contested modes) to help test the feasibility of that argument. This would mean that a mix of formal and informal regulatory initiatives are examined.

Given that all of these 11 potential cases have all been unexplored from a regulatory politics standpoint, I selected two regulatory episodes that were paradigmatic cases that displayed within case-variation and could provide a probability probe for the general arguments advanced in this thesis.

The episode which seems to be the best first-choice case study is Germany and the NetzDG, as it was the world's first regulatory initiative of its type, and one which has

not only had a deep influence on not only the French and Austrian regulations from 2020 that were directly modelled after it, but also, some researchers have argued, provided a global precedent that influenced the decision of other countries to adopt similar rules (W. Schulz, 2019, p. 14). It also provides within case variation, and a regulatory process that led to both collaborative and contested strategies, and an apparent period of evolution between the two. Researching this case is additionally more accessible due to my location in Europe, and my basic German language skills.

The second case study should thus provide a different type of variation. The most attractive choice is one that not only shows evolution between different strategies within a country, but comparatively across two neighbouring countries. Following the Christchurch Attack in spring 2019, Australia and New Zealand both undertook very different regulatory responses to seeking to govern platform practices around violent extremist content, with a collaborative governance mode emerging in New Zealand while the Australian government instead engaged in contestation. This episode shows how two neighbouring countries with strong economic, social, and political linkages pursued the same stated policy goal, following the same event and in the same time period, with two significantly different strategies, thus providing an opportunity to better observe and isolate the factors leading to the change that I am interested in. This regulatory episode provides an additional advantage of featuring primarily English-language documents and reporting.

While in an ideal world it would be possible to look more deeply at other cases, and perhaps a future book project growing out of this thesis will seek to do so, Germany, New Zealand, and Australia provide the best possible set of cases to demonstrate the feasibility of the argument outlined in Chapter 2 of this thesis. Nevertheless, I mention other cases throughout, most notably drawing upon the Singapore POFMA and the French Loi Avia episodes as shadow cases of sorts, helping to show counterfactuals where different choices may have led to different outcomes.

4

Germany: The Development of the Network Enforcement Act (NetzDG)

Contents

4.1 Introduction

On the 1st of January 2018, the 'Network Enforcement Act' (*Netzwerkdurchsetzungsgesetz*, commonly called 'NetzDG' for short) officially went into effect in Germany. The law, which legally enshrined a number of new obligations for the largest platforms for user-generated content, became the first regulation in the world to directly proscribe how platforms moderated harmful content. It involved a number of new rules, establishing background standards for how firms set up their complaints handling procedures, mandating a designated contact point through

which the authorities could channel specific inquiries and complaints, and setting up a level of mandatory transparency reporting for platform content moderation. At the core of the NetzDG was an obligation that companies remove content that was 'manifestly illegal' under a set of provisions in the German Criminal Code within 24 hours of it being notified. These new obligations were underpinned by an enforcement mechanism that threatened fines of up to 50 million Euros in the case of multiple, systemic violations.

The NetzDG immediately became controversial and politicized, with industry and civil society voicing their concerns about the effects on freedom of expression and the possibility that the financial sanctions would incentivize companies to over-remove reported content (W. Schulz, 2018; Tworek and Leerssen, 2019). For commentators, the law highlighted a clash between not only the German and American normative and legal standards around free expression, but also questions about the ability of individual jurisdictions to assert their authority against the global private standards being enacted by multinational platform companies (Claussen, 2018; Echikson and Knodt, 2018). For supporters of the law, the NetzDG was a significant victory, forcing companies to significantly improve their complaints handling mechanisms and increase the number of people that they employed (either directly or through third-party contractors, as content moderators in Germany), thus demonstrating the primacy of German law over unaccountable systems of private rule-making (He, 2020).

Because the NetzDG was the first national law to explicitly move beyond general liability structures and assess exactly how companies conduct content moderation, it is a vital case for understanding how and why platform content regulation occurs. Significant academic and public commentary has discussed the legal and normative aspects of the NetzDG regulation, with much analysis pointing out various potential shortcomings with the NetzDG process and framework (Heldt, 2019; W. Schulz, 2019; Spindler, 2017). However, there have been no efforts to comprehensively assess the political dimensions of the NetzDG, and the regulatory politics that led up to its adoption. This chapter provides such an overview, conceptualizing the negotiation

of the law as a key episode of political contestation over private and public authority in the platform governance domain. Seen through this lens, an underappreciated aspect of the NetzDG comes to the fore: that the policy process that eventually led to the NetzDG going into full force on the 1st of January 2018 actually began in 2015, through a collaborative platform governance initiative organized by the German Ministry of Justice and Consumer Protection in partnership with Facebook. Why did the initial strategy of collaborative platform governance emerge, and then why, less than two years later, did Germany instead seek to contest private platform authority and impose its sovereignty over platform companies by seeking to layer a distinct, national rules-focused implementation infrastructure on top of the existing policies and practices of firms?

In the chapter that follows, I examine the regulatory episode from 2015-2018 that led to the development of the NetzDG, drawing upon 25 interviews with stakeholders involved the law's negotiation, as well as a trove of new primary policy documents obtained via freedom of information requests filed to various branches of the German government and the European Commission.[1] The core of the argument is that, in 2015, following the external shock of the global 'migration crisis' and a domestic political landscape that increasingly appeared to feature far-right extremism, racism, and xenophobia, certain German politicians (most notably the Minister of Justice Heiko Maas) demanded changes in the private platform status quo. However, the combination of relatively low levels of demand with a normative landscape that had interventionist characteristics in general, but was predisposed towards a self-regulatory strategy for online content more specifically, led to a move towards collaborative platform governance via a 'Task Force' and code of conduct negotiated by firms, civil society, and regulators.

Nevertheless, demand for change grew amongst German policymakers in 2016 and 2017. This appears to have partially been due to issues of implementation and enforcement with the code of conduct, leading to a perception in the governing coalition that the platforms were not taking taking these voluntary commitments

[1] See the Methods Appendix A for a full overview of the interview process, as well as a discussion of the FOIA strategy used.

seriously, as well as the external shock of the US election and the global 'fake news' discourse that changed the way that major platform companies were perceived and portrayed. As the German executive increasingly demanded change that would properly contest the status quo of private platform authority, and return foreign companies to what was perceived to be their rightful place as corporate actors bound within the democratically determined context of domestic laws and norms, the German government was able to meet this demand and successfully supply new rules via a contested governance strategy for a few reasons. Firstly, Germany had the regulatory capacity to do so domestically and the power resources to overcome the powerful transnational institutional constraints of EU regulatory harmonization (exerting political pressure and will as the largest and most powerful EU state to prevent the Commission from intervening against rules which were technically against existing EU intermediary liability legislation). Additionally, German policymakers saw the NetzDG as clearly within their normative remit and scope of appropriate policy intervention, drawing upon the *Rechtsstaat* tradition in Germany legal theory to overcome normative concerns about potential repercussions for freedom of expression. The combination of demand for change, the power to intervene, and an interventionist norm all combined in 2017 to yield the world's first law designed to directly contest the authority of platform's private regimes for adjudicating potentially harmful online content and behaviour.

The analysis presented in the chapter provides a number of contributions for both our understanding of the NetzDG as a key regulatory episode more specifically, as well as to the scholarly conversation on the factors that matter in platform regulation more broadly. The chapter highlights the debate and informal politics between policy entrepreneurs seeking domestic change in Germany and a European Commission seeking to maintain regulatory harmonization across the Single Market, showing how Germany managed to layer their new rules (despite the evident legal and constitutional questions those rules raised) in a successful effort to eventually erode a European status quo that key German policy entrepreneurs, like Minister of Justice Heiko Maas, were unsatisfied with. It demonstrates the importance of domestic

demand for new rules (and the ability or inability of key actors to shape that demand by exerting political voice), and of various under-emphasized institutional factors in the platform policy-making process that range from parliamentary procedure to regulatory notification mechanisms. It additionally provides the first yet study of the NetzDG grounded in original qualitative data.

The chapter proceeds in three parts. The first section provides a quick exploration of the cast of characters (the relevant actors involved in this case, and their preferences), an overview of power resources, and a summary of the normative landscape at play. The second section examines the roots of the NetzDG in a 'Task Force' that was set up in 2015 by the Federal Ministry of Justice and Consumer Protection. The final part looks at the drafting and negotiation of the NetzDG law.

4.2 Regulatory Context

4.2.1 Actors & Preferences

Before diving into the regulatory episode of interest, it may be helpful to provide a general overview of the various actors that will be discussed in the rest of the chapter, and to lay out some background context regarding the regulatory capacity and normative landscape at play. There are four groups of actors that are especially important for the case study: the elected executive branch of the German government, the relevant German ministries (especially the Ministry of Justice and Consumer Protection), the European Commission, and the platform companies.

Within Germany, the main political actor is the executive branch of the government, which was from 2013-2017 composed of a 'grand coalition' between the country's two largest federal parties, the centre-left Social Democrats (SPD, 193 seats) and the centre-right Christian Democrats (311 seats including those of its Bavarian counterpart, the CSU, which operates together with the CDU at the federal level). These two parties in government shared ministerial appointments (with SPD Foreign, Justice, and Labour ministers) and also generally vote together as a bloc in parliament. In 2013-2017 there were two opposition parties: the Greens (*Die Grüne,* 63 seats), a socially, economically, and environmentally progressive

party born out of the 1970s student movements; and the Left (*Die Linke,* 64 seats), a left-wing group founded in 2007 with ties to the former governing party of the East German Democratic Republic. Drawing upon scholarship from legislative choice, we can generally assume that the preference of these various elected sub-state actors is not only for re-election (Mayhew, 2004), but also for achieving policy goals in order to signal credible commitments to constituents and to advance their political agenda (Fujimura, 2016).

Additionally, there are a number of regulatory agencies composed of non-elected civil servants that play an important role in drafting, implementing, and enforcing rules. After the scope of a legislation is determined in by the executive branch of government (the Federal Chancellery, and the cabinet) the text of the draft law is written up by officials inside the competent ministry involved in a policy issue. Because Germany has no ministry for digital policy issues, the competent ministry over digital issues is usually the Ministry of Justice and Consumer Protection (BMJV). Issues relevant to security and the economy might also feature the Federal Criminal Police Office (BKA), the Federal Intelligence Service (BND), or the Federal Ministry of Economic Affairs and Energy (BMWI). While these ministries may have their own specific policy agendas, I conceive of their core preference broadly as the supply of rules to meet the demand for rules put upon them by the executive branch.

Because Germany is part of the European Union, a few political actors at the EU level are important as well. While the broad thrust of EU policy is increasingly being dictated by the European Council of (Member State) Ministers, the European Commission maintains the broad policy-making remit for the EU, especially in more technical areas (Rauh, 2019). The Commission is structured into around thirty directorate generals (DGs) which cover various policy areas. On digital policy, the most important ones are DG Communications Networks, Content, and Technology (CNECT), DG Justice and Consumers (JUST), DG Competition (COMP), and DG Internal Market, Industry, Entrepreneurship and Small-to-Medium Enterprises (GROW). While the politics of EU policymaking, and of the Commission in particular, are hugely complex, work has established that

each DG is motivated by its "own competence-seeking motives, varying stakeholder networks, and Commissioners with different national and partisan backgrounds" (Rauh, 2019, p. 352; see also Hartlapp, Metz, and Rauh, 2014). Nevertheless, one can generally assume, following EU regulatory politics scholarship, that the overarching preferences for the Commission and its DGs are the maintenance of a single European Market, and the prevention of regulatory fragmentation across member states (Bradford, 2020; Vogel, 2003).

Finally, the industry that is the regulatory target in question constitutes an important actor group. The platform companies that are subject to new potential rule changes — in this chapter, Facebook, Google, and Twitter — are transnational corporations that all have global headquarters in the Bay Area of San Francisco and their European headquarters in Dublin, Ireland. While there is surprisingly little work on the specific regulatory preferences of platform companies in the global context (although some attention has been paid to the strategies that they use to achieve their aims; see Culpepper and Thelen, 2019), we can extrapolate from the broader literature on corporations and transnational regulation to make a few simple assumptions. Firstly, in global governance scholarship, publicly traded companies are generally assumed to be motivated largely by their need to pursue profits and growth in their quarterly reports to shareholders, although constructivist scholars have contested the universal accuracy of this depiction (Brühl and Hofferberth, 2013). Given that platform companies appear to have derived much of their success and profitability due to a laissez-faire, advantageous regulatory environment in the United States (Kosseff, 2019), the assumption can be made that they prefer the least costly regulatory burden (Cohen, 2019). This generally means that they prefer less-stringent standards or the (laissez-faire) status quo, although in some cases, it is conceivable that firms may prefer higher standards if this means reducing potentially costly regulatory uncertainty in a jurisdiction (Bradford, 2020).

4.2.2 Power Resources & Institutional Constraints

In traditional, power-based terms, Germany would likely be said to wield significant market power. Germany had in 2015 the fourth largest GDP in the world on nominal terms.[2] The largest country by population in Europe, its approximately 80 million residents in 2015 put it within the top twenty of the world's most populous countries. However, the penetration of major platform companies was not as high as in the United States or other leading markets; while precise statistics are difficult to come by, less than 40% of the German population is active on various user generated content platforms. For instance, in 2015 there were only about 25-30 million Facebook users in Germany, making the country roughly their 19th largest market.[3] The best estimates of Twitter usership in Germany put it at about 5 million users in 2021; YouTube has become more popular in the past few years, according to some recent estimates, with 2019 figures suggesting it is used by almost 75% of German social media users.[4] According to some estimates, platform companies like Facebook also glean about five times less revenue per user on average in Europe, as North America is far bigger and more lucrative advertising market, minimizing the policy impact of EU countries on platform companies.[5] Despite the large GDP of Germany, Germany's market power *vis a vis* platform companies might be considered moderate rather than substantial.

That all said, and as outlined in Chapter 2, market size is not the only important factor in shaping the ability of a state actor to intervene and supply new rules. Germany can potentially punch above its market power weight by exerting its considerable influence on the supranational bloc of the European Union, a global regulatory power and frequent exporter of regulatory standards which has its policy preferences shaped by the preferences of leading member states (Bradford, 2012). Germany thus has some influence on shaping regulatory outcomes in the large EU market — which for example, represented approximately 320 million Facebook

[2]See the data from the World Bank, available at `https://data.worldbank.org/country/DE`.
[3]See Statista Research Department (2016).
[4]See Tankovska (2021) and Content Works (2019).
[5]See Tankovska (2021).

users in 2015, a significant proportion of Facebook's approximately 1.45 billion users in the second half of 2015.

Today, Germany is known for having a highly competent and powerful bureaucratic state, with well funded and capable regulators (Müller, 2001) — in a major development from the 1980s and 1990s, where the country did not yet have independent or sophisticated regulatory agencies for many sectors (Bach and Newman, 2007). However, Germany notably does not have a federal ministry tasked with digital policy or digitization issues, nor does it have a federal media regulator, meaning that both the relevant expertise and competencies are diffused across various state level and federal ministries.

Also important to consider for a state's power to intervene are any institutional constraints that are shaping a state's ability to supply new rules. While Germany's role within the EU can provide an amplifying effect, allowing it to potentially steer EU policy and tap into broader EU competencies (and the tens of thousands of regulators working at the European Commission), Germany's membership in the EU also presents it with a number of transnational institutional constraints. The most important one of these is that the German government must generally adhere to existing EU laws and regulatory frameworks when developing new rules. Under a series of measures designed to ensure harmonization of the European single market, codified into European law by Directive 1998/34, and most recently updated in the Single Market Transparency Directive 2015/1535, the EU has a procedure for notification of technical regulations and of rules on products and services, including 'information society services' (e-commerce, media, and internet services). The 2015/1535 Directive sets out a process through which member states must notify the European Commission of any changes to the rules they wish to impose upon certain products or services, including electronic ones, setting up a formalized mechanism through which member states must submit draft laws for review by the Commission and other member states before they are adopted.

The procedure requires a three month 'standstill' period, in which the member state must wait to receive comments from the Commission and other member

states; during this period, the Directorate General for the Internal Market (DG GROW) spearheads a consultation with other DGs, and conducts a legal analysis intended to "help Member States ascertain the degree of compatibility of notified drafts with EU law."[6] This means that an individual member state cannot simply decide to regulate an issue tomorrow, whip up a draft law, and push it through parliament immediately; it must formally notify the Commission (where the draft law is placed in a publicly available database) and wait three months for the input of the Commission and other member states. The Commission, as well as the other Member States, can choose to do nothing, issue a comment to be taken into consideration by the proposing party, or issue a so-called 'detailed opinion': if this occurs, the standstill period is further extended by a month, and the Member State must formally respond to the issues raised by the complainant. Through this notification and harmonization process, the Commission can veto proposed member state draft regulations if it has its own concrete plans to regulate in that area:

> The Commission can block a draft technical regulation if it announces its intention of proposing an EU act (directive, regulation or decision) or its finding that the draft legislation concerns a matter which is covered by a proposal for an EU act presented to the Council. In the case of draft technical regulations containing rules on services, the Commission can block such draft acts only when it announces its finding that the draft legislation concerns a matter which is covered by a proposal for an EU act presented to the Council.[7]

According to the TRIS database, since 1999, 305 national level regulations pertaining to 'information society services' have been notified to the Commission, with detailed opinions issued on around 10 per cent (30) of the proposals. According to the text of EU 2015/1535, any regulatory initiative that changes the rules of operation for businesses — even less formalized processes like voluntary regulations or codes of conduct — should be notified to the commission. If a government fails to notify the law and yet implements it anyway, it can be deemed invalid by the European Court of Justice (see Case C-194/94 CJEU).

[6] A full description is available at `https://ec.europa.eu/growth/tools-databases/tris/en/about-the-20151535/the-aim-of-the-20151535-procedure/`

[7] Ibid, n.p.

4.2.3 Normative Landscape

What is the German attitude to regulation more generally, and relating to channels of information distribution and dissemination more specifically? What is the normative landscape shaping the willingness of policymakers to intervene with rules that might have an impact on free expression?

Generally, Germany is known for being a neo-liberal state with an unusually high, ordoliberal or 'new corporatist' degree of government intervention in markets (Streeck, 2009; Witt, 2002). It has a relatively high appetite for intervention in media, communications, and information sectors generally, although this has been structured in an intentionally decentralized and slightly quixotic fashion. Media policy in Germany is complex, and was born in the aftermath of the Second World War as part of a broader effort to avoid politically dangerous concentrations of power and the likelihood of vital communications infrastructure being captured by anything akin to the Nazi regime. Information and communication policy is thus largely decentralized and placed under the remit of the federal states, as established in the German Interstate Broadcasting Treaty (*Rundfunkstaatsvertrag*, or RStV) of 1991. Beginning in the 1990s, pressure was placed on media companies and the emerging telecommunications industry to embrace certain self-regulatory measures, especially in the realm of youth protection, and Germany has a long tradition of self-regulatory and co-regulatory management of information distribution sectors (Hoffmann-Riem, 2016). The German Association for Voluntary Self-Regulation of Digital Media Service Providers (*Freiwillige Selbstkontrolle Multimedia-Diensteanbieter,* or FSM) was established in 1997 and eventually became a major part of the regulatory framework for youth protection in the media that was codified by the German states in the Interstate Treaty on the Protection of Human Dignity and the Protection of Minors in Broadcasting and in Telemedia (*Jugendmedienschutz-Staatsvertrag*, JMStV) of 2002. FSM works within the tradition of 'regulated self-regulation,' a German regulatory approach which involves self-regulatory associations of companies being overseen by some body that meets criteria set out by regulation (e.g. the institution must meet legal criteria, be licensed by the state and monitored by the

Federal Office of Justice). This approach emerged in the early 2000s and has been a major way through which Germany has sought to regulate new media industries (Hoffmann-Riem, 2016; W. Schulz and Held, 2002). FSM has developed a number of code of conducts for its members, including a code on self-regulation for search engines in 2005 (Google, ask.de, MSN, Searchteq, T-Online, and Yahoo!) and a code for major mobile phone providers in 2007.[8] In 2009, FSM spearheaded a code of conduct with the largest social networks then active in Germany (VZnet Netzwerke, Lokalisten and wer-kennt-wen), all of which would eventually be made irrelevant by the rise of giant American alternatives.

Germany also has a historically distinct position on freedom of expression. In all European Union countries, freedom of expression is encoded under the 1950 European Convention on Human Rights (ECHR) — the first paragraph of Article 10 of the ECHR notes that "Everyone has the right to freedom of expression... This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers" (Benedek and Kettemann, 2014, p. 23). The European human rights treaties also include provisions which describe the conditions under which the right to freedom of opinion and expression can be restricted, which include the prevention of crimes, disorder, and incidents which lead to the violations of the rights of others. Nevertheless, Germany is known for having a more restrictive environment for free expression than many other democracies, even though freedom of expression is enshrined as a fundamental individual right in Article 5(1) of the constitution, the Basic Law (*Grundgesetz*) (Karpen, Molle, and Schwarz, 2007). While the Basic Law notes that "there shall be no censorship," which is understood as restriction on certain types of restrictions rather than a blanket US-style free speech exceptionalism (Jouanjan, 2009), Article 5 also notes how the right to expression may be limited by certain general laws and youth protection statutes.

In the German Criminal Code (*Strafgesetzbuch*, or StGB), a wide array of offenses that disturb the public peace and the concept of the 'free democratic basic

[8] See `https://www.fsm.de/en/voluntary-commitments` for a detailed discussion.

order' are prohibited (Appleman, 1995). Under the StGB, it is famously illegal to disseminate the propaganda or symbols of unconstitutional organizations, such as those associated with the German Nazi Party (Sec. 86 and 86a); to defame the state and its symbols, including its flag, colours, or anthems (Sec. 90a); engage in criminal insult or defamation (Sec. 185 and 186); or to incite hatred against national, racial, or religious groups (Appleman, 1995, p. 413). Evidently, this is a far less absolutist position on free expression than encoded in other legal approaches, and many different types of harmful speech restrictions are legally justifiable in the German context, even when they come into tension with the broader European legal frameworks (ARTICLE 19, 2018).

As Tworek (2021, p. 115) outlines in a historical analysis of the connection between the NetzDG and older German principles of speech regulation, Germany has long tradition of "seeing speech law as a political solution to democratic problems, especially concerns about the inability of citizens to protect themselves from dangerous [material]." In effect, Tworek argues that successive German governments since the Wiemar era have held a generally interventionist position on governing information channels, seeing the creation of rules for books, radio, television, and now, digital media, as part of the core competency of a democratic government. The specific history of Germany, forged out of two world wars, have led to the country to have an interventionist norm, although one which, paired with Germany's specific regulatory tradition, seems to often favour co-regulatory governance solutions.

4.3 The Task Force, 2015-2017

In 2015, a policy process began to unfold that would eventually culminate in the world's first regulation to directly and systemically contest the regimes of private rules governing online speech and activity created and enforced by user-generated content platforms. The following sections provide an analysis of the NetzDG, demonstrating how it evolved from a voluntary 'task force' and code of conduct, into a piece of controversial and politicized statutory regulation.

Dissatisfaction with the Status Quo

In the summer of 2015, in the midst of a major influx of refugees displaced in the Syrian Civil War, Germany, economically and politically the most powerful state in the European Union, decided to break with the established EU resettlement approach under the Dublin Regulation, stating that they would accept asylum claims from Syrians even if their port of entry into Europe was another country (Dernbach, 2015; Hinger, 2016). This policy move was morally laudable but politically controversial, catalyzing far-right extremist groups opposing the re-settlement of refugees in Germany (Dostal, 2015), eventually manifesting in anti-refugee rallies and physical assaults upon immigrants. The number of reported criminal offenses targeting refugee re-settlement facilities would skyrocket from only 24 in 2012 to several hundred in 2015 (Gathmann, 2015), prompting a heated national conversation on immigration, racism, and multiculturalism. German Federal Chancellor Angela Merkel, discussing the situation in August 2015 after visiting a refugee centre in the Eastern state of Saxon-Anhalt, where she had been booed and harassed by right-wing protesters, infamously quipped that despite the challenges, 'we can do it' (*wir schaffen das*) — that Germany was a strong country, had accommodated those fleeing war and persecution in large numbers before, and could do it again.

As the humanitarian crisis unfolded, so did the apparent visibility of far-right extremism and Islamophobia on major social networks. Major figures in German politics, including Merkel herself, were being targeted by online harassment and threats, and commentaries in the country's largest newspapers had begun to point the finger at the content standards on Facebook and Twitter. For example, an emblematic article published in *Der Spiegel*, the weekly news magazine with the largest such circulation in Germany, posed the question of "Why Facebook doesn't delete Hate," bringing up multiple anecdotal instances of public comments left on the Facebook pages of German news outlets not being removed despite being user reports (Reinbold, 2015). The article noted that Facebook was extremely opaque about its content moderation processes — what the exact rules against racist content were, and how those rules were enforced, and by whom — arguing

that the company appeared to conduct moderation via a network of contractors in Dublin, India, and the US, but apparently had no actual content moderators in Germany itself (Reinbold, 2015).

Amidst these external political shocks, according to interviews conducted with policy advisers in Germany's Christian Democrat-Social Democrat (CDU-SPD) grand coalition, dissatisfaction with the status quo began to build amongst key German decision-makers.[9] The most important of these was Heiko Maas, who became the SPD Minister of Justice and Consumer Protection in the CDU-led coalition government formed following the 2013 election. Maas was a vocal critic of right-wing extremism and anti-refugee sentiment, speaking out on numerous occasions against far-right and anti-immigration political movements in 2014 and 2015, and was also an active social media user (Vasagar, 2014). In the summer of 2015, he wrote a letter to Richard Allan, Facebook's head of public policy for Europe, in which he voiced his displeasure with how the company had been handling complaints around illegal or harmful speech, including slurs directed towards refugees and immigrants. Maas noted that "Facebook users are, in particular, complaining increasingly that your company is not effectively stopping racist 'posts' and comments despite their pointing out concrete examples" (Kirschbaum, 2015), in perhaps the first indication that there was significant dissatisfaction with the regulatory status quo at the executive level in the German government. The language of Maas' letter was the public articulation of what some digitally oriented policymakers had been arguing for several years: that the status quo for how major platforms conducted content moderation had shifted away from an 'imperfect but good enough' situation and towards one where the rules were becoming wholly unacceptable.[10]

Negotiating (Voluntary) Commitments

On the 14th of September 2015, Maas met with Facebook's Richard Allan, and at a short press conference that followed, announced that the two had agreed

[9] Interview held via videocall with CDU staffers that requested anonymity, July 2020.
[10] See e.g the writing of von Notz (2015) and others in the German Green Party.

to start negotiating some new measures through a collaborative and voluntary regulatory initiative which would address standards around illegal online hate speech.[11] Through this 'Task Force Against Illegal Online Hate Speech', Maas promised to engage both Facebook and civil society stakeholders in order to produce "concrete measures" for the companies to implement by the end of the year. In an interview that was published a few days later by the *Jüdische Allgemeine*, a newspaper serving the German-Jewish community, Maas delivered a simple message that would become the catch-phrase of the Ministry's agenda to regulate social networks. As he vowed to fight against online anti-Semitism and other forms of platform-mediated hate, he stated simply the Task Force's aim to bring German rule of law to the online, platform-mediated public sphere: "what is forbidden offline is also not allowed online" (Krauss, 2015).

Importantly, in these early days the effort to affect platform content standards was in effect being solely driven by Maas as a policy entrepreneur: Maas was the source of the demand for new rules, which had not appeared to have spread to the rest of the German executive.[12] This level of demand thus remained relatively low, and there appeared to be a clear normative and institutional playbook to try and meet this demand: the tradition of 'regulated self-regulation' in Germany, where task forces and codes of conduct developed collaboratively and overseen by government had been deployed in other telecommunications and media industries, and was the status quo for interventions into areas like online content on search engines (Hoffmann-Riem, 2016; W. Schulz and Held, 2002). At this point, one might argue that Maas did not fully yet have the adequate demand or power to intervene without mobilizing the rest of the executive, making the collaborative approach the most natural option, one that fit normatively within the German institutional tradition of content regulation (He, 2020).

[11] Maas' statement to begin the press conference is available online at: `https://www.youtube.com/watch?v=TZdWdrfDnug&feature=youtu.be`

[12] Interview with CDU staffers; and interview held via Signal with Joern Pohl, Chief of Staff, MdB Konstantin von Notz, May 2020.

The Task Force had its first meeting ten days later in Berlin. Following an opening by Gerd Billen, the most senior civil servant in the Ministry of Justice, who had been tapped by Maas to lead the Task Force, the meeting featured presentations from Facebook and Google and concluded with inputs from the handful of civil society and hybrid civil society/governmental organizations that attended.[13] At the onset, very little public information was released about the project, and detailed meeting minutes were not kept.[14] The task force had 4 meetings in 2015: September 25, October 10, December 7, and December 15, with the participants including representatives from Facebook, YouTube, and Twitter; the industry associations eco and FSM; and four German organizations working on issues relating to child protection, racism and far-right extremism.[15] Together, the participants in the working group began negotiating a possible set of commitments, with Ministry officials pushing for content reported in Germany to be reviewed in Germany and for broader application of German law rather than company community standards.[16]

On December 15, after a 4th meeting of the group, a 5-page 'results paper' (*ergebnispapier)* from the Task Force was published. This document sets out the "concrete measures" that Maas had promised by the end of the year when announcing the initiative, and in it, the companies make a number of commitments to improve their standards for complaints processing "by mid-2016." This document does not explicitly refer to itself as a code of conduct, but in interviews I conducted with

[13]Partial agenda summaries are available on an archived Ministry webpage: `https://web.archive.org/web/20170930061101/http://www.fair-im-netz.de/WebS/NHS/DE/Home/home_node.html`

[14]In a freedom of information request to the Ministry, to obtain the meeting minutes for the Task Force, the ministry responded that they did not exist (Fraag den Staat, 2017)

It seems as if the BMJV created a website (no-longer online) which had more details about the task force, with brief summaries of the meetings and the main commitments made, but this was only archived in July 2017 (suggesting it was created during that key legislative moment and then later taken down. The last archive was in July 2019.) `https://web.archive.org/web/20170930061101/http://www.fair-im-netz.de/WebS/NHS/DE/Home/home_node.html`

[15]These were jugendschutz.net, klicksafe.de, the Amadeu Antonio-Stiftung, and Gesicht Zeigen. While the Amadeu Antonio foundation and Gesicht Zeigen are independent civil society organizations, Jungendschutz and Klicksafe are probably better understood as governmental actors or quasi-governmental actors with close ties to the German state and federal governments. Klicksafe is a EU funded project of the state media regulators of Rhineland-Palatinate and North Rhine-Westphalia.)

[16]Interview held via videoconference with Simone Rafael, Executive Director of the Amadeu Antonio-Stiftung, June 2020.

individuals who attended the Task Force's meetings, the interviewees repeatedly referred to a "code of conduct" as the central result of the Task Force.[17] The main take-away of the document is its emphasis the companies will act against "all hate speech prohibited against German law" and "review and remove without delay upon notification".[18] To achieve that goal, the document outlines a few 'best practices' and other commitments that have varying levels of clarity and ambiguity. The three main parts of these commitments are published in an infographic that Maas shares on Twitter, which summarizes the code of conduct for the public as follows: companies (a) agree to respect German law (in other words, what is illegal offline should be illegal online), (b) agree to remove reported content in less than 24 hours, and to (c) improve their user-reporting tools.[19]

No formal institutional structures had been changed, but a new set of general and voluntary rules had been layered on top of firm's existing commitments under EU and German law; additionally, the Task Force created a forum for information sharing, negotiation, and discussion between firm policy employees (importantly, not firm *policymakers*, however; it does not appear to have been the case that the higher level platform employees with the power to actually make meaningful changes to firm rules and processes attended Task Force meetings — instead, lower ranking regional or German representatives attended). The companies agreed to implement the terms of the code of conduct in the next six months, but this was an informal agreement, and the publicly-released results paper was not undersigned by the companies or specific employees.

[17]Interview conducted with Simone Rafael, Antonio Amadeu Stiftung; and via video conference in April 2020 with Lutz Mache, Public Policy and Government Relations Manager, Google Germany.

[18]These quotes are from p. 1 of Ministry's official English translation, obtained by EDRI and available here: `https://edri.org/eu-internet-forum-document-pool/`

The German version is archived here: `https://perma.cc/J35T-DGC6`

[19]The tweet is available at `https://twitter.com/HeikoMaas/status/676739434239426561`

4.4 The Network Enforcement Act, 2017-2018

Demand for Changing the (Collaborative) Status Quo

On the 11th of April 2016, press releases from the German Ministry for Family Affairs and the Ministry of Justice announced that the two ministries would be working together to commission a monitoring exercise to evaluate the effects of the Task Force's voluntary commitments (BMFSFJ, 2016). This informal evaluation would be performed by Jungendschutz.net, an organization that was established in 1997 with funding from the Ministry of Family Affairs and serves as a 'centre of competence' for the German states on child protection issues. Since 2008, Jungendschutz has been conducting research and advocacy into online child safety, with their legal mandate set out in the Interstate Treaty on the Protection of Minors (JMStV). Beyond actively searching out illegal content and reporting it to the platforms (in their 2008 annual report, for instance, they claim that they successfully were able to secure the removal of 1400 illegal videos from YouTube either in Germany or globally), they had from 2008 onwards conducted a number of simple audit studies, in which their employees would proactively attempt to find illegal content on search engines or social networks (Glaser et al., 2008). Through a collaboration with the Ministry of Justice, Jungendschutz brought some research capacity and expertise when it came to content moderation standards, even though their thematic focus was on a different issue area (child protection, not hate speech). As the BMJV's Gerd Billen noted in a statement, the monitoring would be an "important component of the task force:"

> The monitoring provides us with important insights into how agreements with companies work in practice, how quickly they react to reports and whether they delete the reported illegal hate content. This will enable us to better assess how the agreed measures are taking effect and what further steps are necessary (BMFSFJ, 2016, np, author translation).

Jungendschutz employees conducted their first formal evaluation in July 2016, by which point the firms were supposed to have implemented the code's commitments (Jungendschutz, 2016). The results were not in line with the Ministry's expectations.

As Maas later summarized at a public event, the figures released by Jungendschutz, based on a small sample of content takedown requests, suggested that "of the illegal content reported by users, Twitter deletes about 1%, YouTube just 10%, and Facebook about 46%" (Reuters, 2016). Shortly following the evaluation, Maas wrote again to Richard Allan and to Facebook's head lobbyist in Berlin. In the letter, obtained by a freedom of information request, Maas wrote that "the results of your efforts thus far have fallen short of what we agreed on together in the Task Force" (Beckedahl, 2016, author translation). In full awareness that the Task Force commitments were voluntary, and thus there were no sanctioning mechanisms or enforcement capabilities built in, he threatened action at the European level if Facebook did not step up their game — writing that he had been discussing the issue with other Justice Ministers in the European Council and that they 'shared his concerns,' suggesting that he would seek to influence his European counterparts towards pursuing harder and costlier forms of regulation at the European level. (Despite the even poorer performance displayed by Google and Twitter on those same metrics collated by Jungendshutz, it does not appear that similar letters were sent to Google or Twitter representatives).

Demand for changing the rules was growing within the German government due to the perceived internal failure of the firms to take the collaborative approach seriously, as well as external, global developments that were increasing the salience of platform governance as a transnational digital policy issue. First, as Tworek (2021) and others have noted, domestic legal developments were leading lawmakers in CDU/SPD governing coalition to follow in Maas's footsteps and worry that "German law could no longer be enforced in Germany" on platforms due to jurisdictional issues:

> Amongst several cases filed, one German lawyer, Chan-jo Jun, had filed a case against Facebook for not removing online content that was illegal under German law. In 2016, a regional court in Hamburg denied the complaint on the grounds that it did not have jurisdiction to adjudicate because Facebook's European operations are headquartered in Ireland. Jun called it "outlandish" that American companies could operate in Germany without being subject to its jurisdiction (Tworek, 2021, p. 110).

In multiple interviews I undertook with lawmakers and their staff involved in the policy debate at the time, deep frustration was expressed about the opacity of the companies (especially Facebook) and their unwillingness and/or inability to speak candidly about how they enforced their global content moderation rules in the German context.[20] When firm representatives offered testimony to parliamentary committees or at public events, they refused to provide what was perceived to be basic detail about the number of German-speaking content moderators that they employed and their specific capacities in Germany. (At the time, the firms were extremely cagey about who and how these processes functioned; as Gillespie (2018) has noted, and this served as a strategy to avoid scrutiny and downplay the importance of their moderation practices). But this strategy appeared to backfire in the German context. Despite the measures being instituted voluntarily by the companies through the task force, the perception amongst key stakeholders in German executive was increasingly that the firms were merely doing "whatever they could to avoid regulation totally and limit their costs," as one Member of the Bundestag in the governing coalition put it.[21] As one staffer, the digital policy adviser to a Member of the Bundestag on the Digital Agenda committee noted, throughout the Task Force process, and the effort by German officials to achieve their preferences via collaborative means, "Facebook in effect told [German lawmakers] to their faces that 'yes, the issue is complicated, but we're sorry, but we can't accept your national [criminal] laws.' "

Domestic demand for higher standards for platform content moderation practices was also spreading to the rest of the executive due to the exogenous shock of the election of Donald Trump to the US Presidency in November 2016, following a scandal-filled and salacious campaign where social media platforms, foreign interference, and the influence of 'fake news' were all said to have played an outsized role (Karpf, 2017). Following a strong performance by the far-right Alternative

[20]Interview held via videoconference with Member of the Bundestag (2013—) Jens Zimmerman, SPD Digital Policy Spokesman, June 2020; interview held via videoconference with Alexander Ritzmann, Counter Extremism Project, June 2020.

[21]Interview with MdB Jens Zimmerman.

for Germany (AfD) party in a number of 2016 German state elections, where the AfD, in a number of cases, appeared to take votes from both the CDU and the SPD, concern was mounting in the governing coalition that various forms of digital trickery could have an adverse effect on their electoral outcome in the German Federal election that would be happening in Fall 2017.[22] As Gollatz and Jenner (2018) have documented via qualitative and quantitative media analysis, the post-US election's 'fake news' discourse quickly the domestic German debate on the NetzDG and helped to frame it as a threat to the democratic integrity of Germany in the context of the upcoming election.

On March 14 2017, Heiko Maas publicly announced a new draft law designed to layer platform-specific content moderation rules for firms on top of the existing European intermediary liability regime. The law, the *Gesetzes zur Verbesserung der Rechtsdurchsetzungin sozialen Netzwerken* (literally, the legislation to improve law enforcement in social networks, commonly shortened to *Netzwerkdurchsetzungsgesetz*; this translates to the 'Network Enforcement Act' in English, with the abbreviation *NetzDG* commonly used in both German and English-language writing about the regulation), was a 29 page piece of draft legislation with a number of obligations set out for the regulatory targets:[23] they would have to (a) publish quarterly reports on the handling of complaints about illegal content made by users; (b) delete 'obviously illegal content' within 24 hours, and other forms of illegal content within 7 days; (c) appoint a contact person to receive government queries and complaints in Germany; (d) inform users about content moderation procedures through various means; (e) save or archive content removed as illegal for prosecutors to use as evidence, and (f) search for and delete copies of illegal content that existed in other places on the platform. The preface to the draft law outlined the estimated costs that would be

[22]Interview held via videoconference with Wolfgang Schulz, Director of the Alexander von Humboldt Institute for Internet and Society, April 2020.

[23]These targets were defined broadly as 'service providers who operate platforms on the Internet with the intention of making a profit which enable users to exchange, share or make available to the public any content with other users (social networks)', with the exclusion of journalistically curated services where an editor is responsible for content under existing legal frameworks (e.g. newspapers, broadcasters).

The first draft is available at: `https://cdn.netzpolitik.org/wp-upload/2017/03/1703014_NetzwerkDurchsetzungsG.pdf`, author translation.

an outcome of the regulation: approximately 28 million Euros in annual compliance costs for all firms, reflecting increased staffing costs and the cost of putting together the transparency reports, and an estimated 3.7 million Euros annually in terms of bureaucratic costs for the government.

Attempts to Tamper Down Demand

Immediately following the announcement, there was a strong backlash from digital civil society, as well as from global human rights organizations and from industry lobby groups seeking to suppress this demand. A number of civil society organizations, including the German digital rights organization *Digitale Gesellschaft* and the global press freedom organization Reporters Without Borders likewise predicted that the law would have deleterious effects on freedom of expression. D64, a network of digital policy experts that is closely linked to the SPD, called the law (and specifically, it's provision for the automatic deletion of matched content) "the first step in a creation of a censorship infrastructure" (Reuter, 2017, author translation). The UN Special Rapporteur for freedom of opinion and expression noted in a letter to the German executive branch that the law would incentivize the overblocking of legitimate speech by users, as it was formulated around the main metrics of 'takedowns' and 'speed,' with no real mechanism for auditing the rates of false-positives made by the companies (Kaye, 2019). A coalition of both transnational civil society organizations as well as industry groups published an open letter against the NetzDG, eventually securing a series of high-level meetings with lawmakers in the governing coalition to try and negotiate concessions or its withdrawal (He, 2020; Reuter, 2017).

Industry was unsurprisingly also strongly opposed to the proposal: Bitkom, a industry lobby group that counts Facebook, Twitter, and Google amongst its members, immediately issued a statement warning that the law would spur a "takedown orgy" (*Löschorgien*) as firms would be incentivized to over-remove content rather than face fines for acting too slowly (Reuter, 2017). Other groups warned that the law, as perhaps the first in the world to regulate content moderation as done by

platforms in a non-copyright and intellectual property context, and additionally one being proposed by such an internationally influential and democratically legitimate state like Germany, would serve as a model for other less-democratic governments seeking to bring social media companies under closer state control (He, 2020; Tworek, 2021). The furor was intense and as the critiques in major media outlets circulated widely, Maas had to answer the critics in a number of interviews in mid-March with major outlets like *Der Spiegel* and in debates with civil society (Gathmann and Knaup, 2017).

Nevertheless, lawmakers in the governing coalition were supportive of the bill, downplaying the risks and emphasizing the importance of taking a strong position in fighting against illegal content online. The bill appeared to have significant support in the governing CDU/CDU coalition; the legal policy spokespersons for the Union parliamentary group went as far as to say that "The bill by Minister of Justice Maas is a first, small step in the right direction. But we must go much further," suggesting that other types of criminal law enforcement could be also included (Reuter, 2017). A few voices in the coalition expressed opposition — parliamentarians active on digital issues, who had ties to the digital civil society organizations and had above average knowledge on digital policy issues, tended to share at least some of the publicly articulated reservations[24] — but these dissidents were not in major positions within the party and tended to not voice these concerns publicly. Overall, it seemed as if the governing coalition had the political will to keep demand for the new rules high, especially given that they wished to signal their opposition of online hate in the lead up to the election in the fall.

Supply Factors

By this time in 2017, the NetzDG was being proposed as part of a whole-of-government strategy against both online hate and the unaccountable private decision-making power of the US platform companies (He, 2020). Demand for change was high enough that Maas had the blessing of the executive to propose a contested

[24]Interview held via videoconference with Member of the Bundestag (2017—) Mario Brandenburg, Free Democratic Party of Germany, June 2020.

approach with binding rules. He now had the regulatory capacity to develop regulation, tasking a group of civil servants at the Ministry of Justice and Consumer Protection to draft the law, building largely upon the general framework that had been developed through the collaborative Task Force, including the notion that there should be a contact person to handle official complaints, that content moderation standards and complaints procedures should be transparent enough to be clear to users, and that they should generally act on content within 24 hours of it being reported. However, there were also a number of new and quite aggressive provisions proposed in the draft, which appeared to correspond to the new levels of demand in the governing coalition for stronger standards, as well as the more confrontational stance being taken by the German government as far as private platform authority went. These provisions included an obligation for firms to delete duplicates of the content that was deemed manifestly illegal under the NetzDG across their broader platforms (in effect searching for copies of content found to be illegal that had not yet been reported), and a requirement that this deleted content would be archived for potential access by federal prosecutors seeking to bring charges against individuals.

As Maas now led this whole-of-government approach, he could be confident that Germany had regulatory capacity required to intervene and contest private platform authority. His government had a strong majority in the Bundestag, with little that the opposition could do to contest any proposed legislation. While Germany did not have a digital ministry that could have developed more technically sophisticated rules, his civil servants, including Gerd Billen, had been engaged in dialogue with platform companies for almost two years via the Task Force and adjacent informal fora. Their expertise and understanding of the theory and practice of how Facebook, Twitter, and YouTube moderated the content of German users had improved as a result of this direct engagement and capacity building efforts, at least according to the firms.[25]

While the normative landscape shaping the ability and willingness of German policymakers to intervene in this area was of course complex, it appeared to be

[25]Interview with Lutz Mache, Google; interview held via videoconference with anonymous Facebook policy staffer.

shifting towards a space which would allow the creation of binding rules. As a comprehensive analysis by He (2020) has documented, Maas, his Ministry, and others in the German executive were able to increasingly frame the process of the NetzDG as part of the 'Rechtsstaat,' a uniquely German conception of the rule of law that is enshrined in the country's constitution. The concept in effect is that the continued functioning of the German state is dependent on the "existence, validity, and primacy of law... Law can take the form of legislative acts such as the NetzDG, but also, more importantly, of the German constitution, the Basic Law.... The precedence and supremacy of the basic rights enshrined in the Basic Law are crucial to the Rechtsstaat idea" (He, 2020, p. 27). As He argues, Maas was the main policy entrepreneur promoting this idea via his speeches and media appearances, and this notion eventually became the central publicly articulated rationale for the legitimacy and necessity of contesting private platform authority. The debate turned upon the problematic nature of having foreign corporate entities in effect making unaccountable decisions about the speech and behaviour of German citizens at critical political junctures; the NetzDG was positioned as the answer to that problem and a reassertion of public, democratic authority. Although normatively the NetzDG may have displayed a break from the relatively laissez-faire German tradition of online content regulation (which was generally predisposed towards collaborative, self- or co-regulatory arrangements in the few areas it had touched, such as online search engines), it was able to harness a broader interventionist norm in free expression more broadly. As Tworek (2021) has argued in a historical analysis of the NetzDG's normative roots, Germany has long had a far more interventionist conception of the appropriate scope of policy influence in public expression, going back to even Wiemar Germany. The combination of this relatively interventionist normative foundation, when combined with the strategy of discursive legitimization around the 'Rechtsstaat' seemed to put Germany in a strong position to be able to supply binding rules to contest private platform authority.

Transnational Institutional Constraints: EU Harmonization Procedures

Nevertheless, Germany's power to intervene, as discussed in Chapter 2 of this thesis, should according to my conceptual framework also be influenced by various potential institutional constraints. In the NetzDG context, the most significant of these was Germany's role within the regulatory environment of the European Union's Single Market. As the NetzDG was introduced, the European Union's governing institutions were watching closely and deciding exactly what their position should be.

When the German Ministry of Justice and Consumer Protection notified the first draft of the NetzDG through the Technical Regulations Information System on March 27th, that notification was flagged as potentially politically sensitive, according to Commission emails obtained via freedom of information requests. An internal email from a staffer in the Directorate General (DG) for the Internal Market (GROW) — the entity in-charge of managing the TRIS notification system — to other GROW staffers, including members in the office of Internal Market Commissioner Elżbieta Bieńkowska, summarized the main points of the NetzDG, the timeline for reactions from the DGs for Justice (JUST) and for Communications, Content, and Technology (CNECT) and the legal deadline for the Commission or Member States to react. The summary of the notification also discusses the political context, and notes that CNECT is keeping the option of vetoing the law on the table, which can be done if the Commission wishes to pursue a different regulatory strategy:

> The German intention to regulate the matter has been recently discussed between CNECT's Cabinet and the German authorities. During these discussions, CNECT informed the German authorities of CNECT's intention to regulate the same matter with a different approach than the one presented in the notified draft. It seems that DG CNECT and DG JUST are in contact to discuss the notified draft and have contacts with the German Ministry of Justice (which prepared the notified draft).[26]

The Commission had before the NetzDG taken the position that no legal framework for raising content moderation standards for major social media platforms

[26]Gorwa FOIA to DG GROW, 2020. Document received June 9, 2020; document dated April 3, 2017. Available at `https://www.asktheeu.org/en/request/7872/response/26398/attach/3/Document%205.pdf?cookie_passthrough=1`

was necessary. DG JUST had negotiated the Code of Conduct on Hate Speech with the major internet companies in the Spring of 2016, and Vera Jourova, the EU Justice Commissioner, was publicly a major advocate for voluntary self-regulation and co-regulation in areas that would have a major impact on free expression and other fundamental rights. In an internal assessment prepared by DG CNECT and JUST which analyzed the NetzDG and contextualized it within previous Commission measures, Commission staff note that the spirit of the proposal was not totally out of line with their efforts to increase transparency for company content moderation systems and move their private law into a space that it more adequately reflected European legal frameworks:

> While, unlike the [EU Hate Speech] Code of Conduct, the draft German law is a legal instrument, an analysis of its objectives against the objectives pursued in the Code of Conduct shows that the two are broadly coherent in terms of the overall objective. Both instruments aim at ensuring that notifications of illegal hate speech are assessed against the law and not only against the internal terms of service of the IT companies and that the assessment is made expeditiously. An important difference is that the scope of application of the German law goes beyond the Code of Conduct in so far that it includes also other offences, such as defamation.[27]

Nevertheless, the analysis notes that the NetzDG threatens regulatory harmonization as outlined in the Juncker Commission's Digital Single Market Strategy: "The Commission considers that national solutions at this respect can lead to unwanted legal fragmentation and have a negative effect on innovation."[28]

It was clear to officials working in the Commission that the NetzDG was on shaky legal footing. It quite clearly ran against the country of origin principle established in Article 3 of the E-Commerce Directive, which states that Member States may not "restrict the freedom to provide information society services from another Member State" (Hellner, 2004, p. 9),[29]and also clearly had issues on free expression grounds

[27]Gorwa FOIA to DG JUST, 2020. Document received June 16, 2020, and dated June 8, 2017. `https://www.asktheeu.org/en/request/member_state_comments_on_netzdg#incoming-26570`, p. 5

[28]Ibid.

[29]Article 3 of the ECD is legally complex and has been interpreted slightly differently by various member states in their implementations of the ECD. See Hellner (2004) for a detailed discussion.

with European Human Rights law as set out under the European Convention on Human Rights and other measures. As a Commission official involved in the debates at the time discussed, "it was obvious to everyone who had been following the debates in Germany that NetzDG had major issues under European law."[30] However, the situation was just ambiguous enough that what the Commission would do was a political, and not purely legal question. As the official explained, notifying a new law triggered an informal political and legal assessment, and not a fundamental rights compliance assessment, which would only be triggered in the case of the notified proposal transposing European Law (for example, in the case of an amendment to the Telemediengesetz, the German transposition of the E-Commerce Directive). The stakes were high: as one staffer for a Member of the European Parliament working on digital policy issues at the time put it, "the consensus was that early law made by a major member state could serve as a blueprint for eventual European wide legislation."[31]

In effect, the Commission had three formal options. It could issue a comment, a non-binding public response which would advise the German Ministry on changes that the Commission recommended, it could issue a so-called 'detailed opinion' similar to a comment, except one which mandated a reply from the German government and had the additional effect of extending the standstill period by at least a month, or it could try and negotiate these issues off the record in direct negotiations. Because of the timing of the German notification, a detailed opinion would extend the standstill into the Bundestag's summer vacation, and thus past the last session of parliament, effectively killing the proposal.

This made the EU TRIS process an important veto point (Thelen and Steinmo, 1992, p. 7), and the firm and NGO actors that were opposed to the law sought to influence the EU Commission's decision as well, with many of the civil society groups active domestically in Germany writing public comments on the law via

[30]Interview held via videoconference with Prabhat Agarwal, Head of Unit for Online Platforms, DG Connect, May 2020.

[31]Interview held via Signal with Mathias Schindler, Office of Member of the European Parliament Julia Reda (The Greens/European Free Alliance), April 2020.

the TRIS portal. They were joined by transnational civil society networks, as well as industry. In a meeting with DG GROW's cabinet on June 12, Facebook's lead Brussels lobbyist argued that the NetzDG violated the E-Commerce Directive and sought for the Commission to engage in "a dialogue with the German authorities to change the law."[32] A scene-setter with talking points prepared for the Commission official leading the meeting outlines DG GROW's position on the burning question that Facebook was guaranteed to ask: "Does the EC intend to object to the notified German draft"? (The talking point demurs, noting that "The commission is still assessing the compatibility of the Draft Act notified by Germany with EU Law. The deadline for reaction expires... on 28 June 2017").

Overcoming Those Constraints: Dealmaking with the Commission

On the floor of parliament in early May 2019, Maas defended his bill, arguing that it would not lead to privatized enforcement but rather simply to the better existing implementation of German criminal law. Members of the opposition noted that the list of criminal code statutes covered in the NetzDG were extremely broad and went beyond just hate speech (more than 20, including not just incitement to violence and the promotion of unconstitutional organizations, but also defamation and some oddities like the disparagement of the ceremonial President of the Federal Republic), and that the definition of social networks provided in the bill would likely encompass many other online services, like blogs, third-party reviewing sites, and messaging services like Whatsapp or Telegram.[33] Multiple MdBs in both the governing coalition and the opposition complained about the very short time period in which the law was being debated. As Netzpolitik's Markus Reuter observed, "all of the CDU/CSU speakers complained about the little time remaining until summer break" as they proposed their suggestions for changes, including a bigger role for some kind of self-regulatory body used in the media industry to adjudicate

[32]Gorwa FOIA to DG GROW, 2020. Document received June 9, 2020; document dated July 6, 2017. https://www.asktheeu.org/en/request/member_state_comments_on_netzdg#incoming-26398

[33]See contributions from Petra Sitte (Die Linke) and Konstantin von Notz collected by Reuter (2017)

on complaints, rather than the platforms themselves (Reuter, 2017). Similarly, MdB Petra Sitte of *Die Linke* argued in her remarks that given the "broad alliance of organisations that had already formed against the draft law" the governing coalition should "engage in a broad discussion" and revisit the issue following the election to prepare a better proposal (Deutscher Bundestag, 2017).

Overall, the attempts to tamper down demand to head off the NetzDG were ultimately unsuccessful. Domestically, the opposition did not have enough seats to demand concessions or prevent the law's passing. Civil society and firms were unable to exert enough voice against the law to capture regulators or change their preferences, likely due to a combination of the very high levels of demand amongst key policy entrepreneurs, and the apparent failure of firms and civil society to successfully deploy public-oriented campaigns to sufficiently mobilize German platform consumers against the law (as conceptualized by Culpepper and Thelen, 2019). While there is a dearth of good polling data about the NetzDG, the one existing poll with a purportedly representative sample of German social media users (albeit with only a sample size of 500), conducted in early 2018, found that 67% of those polled 'strongly approved' of the policy and 20% 'somewhat approved,' with only 5% of respondents 'disapproving' of the NetzDG.[34] In an interview, an employee of one of the major platform companies suggested that one of the main reasons that their company had not mobilized more aggressively against the NetzDG (in terms of both direct lobbying in Berlin, and in terms of public-oriented PR campaigns) was because internally commissioned survey and focus group research had shown the policy's relatively broad support with the German public.[35]

Additionally, the potential game-changer — the mobilization of US government pressure against Germany to reduce demand, via diplomatic pressure, backroom negotiations, or potentially the threat of retaliatory sanctions against German national champions — never materialized. This may have been due to the tension (and at times overt hostility) between the Silicon Valley giants and the newly

[34]See Jacob (2018).

[35]Interview held via videoconference with major platform policy manager, summer 2020. They requested that both their name and the name of their employer be anonymized.

elected Trump administration, which represented a break from the relatively close relationship with government that the firms enjoyed during the Obama years (Powers and Jablonski, 2015). It also may have simply been that the compliance costs of the NetzDG were not existential enough for the firms to see the expenditure of political capital in Washington to try and get US intervention on their behalf as necessary. In effect, the lack of US government opposition meant that the only real constraint upon Germany's ability to intervene was in Brussels.

The Commission recognized this, and on May 23, a high-level meeting about the NetzDG happened with members of the cabinet for Commissioners Ansip (DG CNECT), Jourova (DG JUST), Timmermans (Commission Vice President), and Juncker (Commission President). As an internal emailed summary of that meeting discusses, Ansip, who was in charge of maintaining the Digital Single Market, "wished to send a political letter to DE on the main concerns [CNECT] have on the draft law," but Juncker, Timmermans, and Jourova did not want to co-sign it.[36] While Ansip argued on the side of maintaining harmonization and using the notification process to get Germany to stand down, the others were hesitant due to a number of political factors. The main one articulated in interviews with officials present at these discussions was that Germany was entering an election year, and there was a perception that if the Commission stepped in and deemed the NetzDG in violation of EU law it could perhaps be perceived as a high profile domestic political defeat for Maas and the SPD, potentially affecting the electoral outcome in some way. A second issue was the culturally sensitive context of hate speech in Germany, and the significant pressure coming from German policymakers, including prominent German staffers in the European Commission, that this issue should be left aside a domestic political matter.[37] Finally, while the DGs may have wished to regulate the issue of harmful content in a different manner than Germany (and indeed, Commissioner Jourova had spearheaded the development

[36]Gorwa FOIA to DG CNECT, 2020. Document received July 10, 2020, and dated May 24, 2017. https://www.asktheeu.org/en/request/member_state_comments_on_netzdg#incoming-27094

[37]Interview held via videoconference with Paul Nemitz, Director for Fundamental Rights, DG Justice, June 2020.

of the EU Code of Conduct on Online Hate Speech in mid-2016), the Commission had no viable policy currently in the works that it could propose to Germany as an alternative to the NetzDG, other than the code of conduct, which followed a similar collaborative approach as the Task Force that was already seen as ineffectual by the German negotiating team.[38]

In a turn from regular procedure and into the realm of informal governance (Kleine, 2013); rather than issuing public comments, the Commission raised concerns through informal letters and other back-channels that would minimize domestic fallback for a German government dead-set on passing the law before the election. This back and forth negotiation, underpinned with the threat of a Commission detailed opinion (and *de facto* veto) successfully negotiated last-minute softening of some of the NetzDG's provisions.

On June 27 2016, the grand coalition introduced an amended version of the bill, "revised in consultation with the Commission in order to achieve the greatest degree of compatibility with EU law" into the Legal Affairs committee.[39] (The language itself on the 'greatest degree of compatibility' possible, rather than actual compatibility, is insightful.) Firstly, the scope of the law was changed slightly, by narrowing the definition of social networks so that it excluded peer-to-peer messaging services, music services, blogs, and other platforms. Combined with a threshold of 2 million registered users in Germany, the law was thus changed so that it would at its onset only apply to Twitter, Facebook, and YouTube (TikTok would find itself in scope in 2020). Secondly, the list of sections of the German Criminal Code that companies would need to check flagged content against was trimmed down, removing a few statutes that had been critiqued by civil society as being redundant and not pertaining to hate speech (e.g. the statues referring to defamation of the Federal President or the 'denigration of constitutional organs' like the courts). Additionally, the new version removed two of the major provisions

[38]Interview w/ Prabhat Agarwal, DG CNECT.

[39]See Gorwa FOIA to DG GROW, 2020. Germany's Response to Sweden, recieved June 9, 2020; document dated July 28, 2017. `https://www.asktheeu.org/en/request/7872/response/26398/attach/2/Document%203%20EN.pdf?cookie_passthrough=1`

that had been added post-Task Force: the provision that firms should have a 'stay down' filter by which they would algorithmically search for, and remove, duplicate content from their platforms when removing an image or video for violating one of the criminal statutes specified in NetzDG, and the provision that the companies would need to archive content deleted for federal prosecutors, which critics argued was especially problematic from a data protection point of view (Reuter, 2017). Finally, the new version added a provision which allowed for the role of 'regulated self-regulation' through voluntary industry bodies to be involved in the reporting or assessment of cases of illegal content reported by users, a prospective safeguard that had been advocated for by the CDU faction.

The Legal Affairs parliamentary committee agreed with this new version, and set a date for the second reading of the bill in the Bundestag three days later, June 30th, on the last day the Bundestag was in session before the election. On the 30th of June, the NetzDG was debated for 45 minutes. The law passed through easily, with the votes of the majority coalition (CDU/CSU and SPD) in favour, with the Left voting against and the Greens in abstention. After receiving official sign-off from Germany's (largely ceremonial) President a few days later, the NetzDG was officially passed into law.

In the days following the vote, it was evident that many lawmakers, even within the grand coalition, were not entirely thrilled with the law, but nevertheless maintained that it was the best that could be done to fill an important policy vacuum given the institutional constraints at play. In an interview with the left-wing daily *Tageszeitung*, the SPD's legal policy spokesperson Johannes Fechner argued that the NetzDG supplied imperfect rules to meet what was a crucially important demand, noting that while the SPD wished to have added more provisions that would have better protected the freedom of expression of users,

> But if we had included a new obligation for companies in the law, we would have had to re-notify the law to the EU. We would then have had to wait another three months to find out whether there were concerns on the part of the EU Commission or other EU states. So the law could not have been passed in this legislative period (Rath, 2017, author translation).

Fechner's comment highlights a number of key points currently missing from the current scholarly and policy conversation on platform regulation, as well as the more narrow discussion around the NetzDG. It demonstrates the lock-in effects that were a result of institutional factors: the EU's notification process and the short time period in which Germany sought to contest the regulatory status quo. Interestingly, the main concessions and changes to the substance of the law were made where the veto points were: in this case, not at the domestic level, where virtually no changes to the NetzDG were successfully negotiated by the opposition parties or by firms and civil society, but at the EU level, in negotiations with the European Commission.

4.5 Conclusion

A detailed policy and process-oriented look at the NetzDG and its origins provides an opportunity for deeper analysis of the conditions under which governments are able to contest platform authority and standards. The core of the conceptual argument laid out in this chapter is that looking more closely at three broad policy factors — demand for change, a state's power to intervene and supply new rules, and the normative landscape regarding the appropriateness of the state in providing those rules — can help us understand how and why certain platform governance modes emerge, and under which conditions they should be expected to be successful.

In 2015, following the external shock of the global 'migration crisis' and a domestic political landscape that increasingly appeared to feature far-right extremism, racism, and xenophobia, certain German politicians (most notably the Minister of Justice Heiko Maas) demanded changes in the private platform status quo. However, the combination of relatively low levels of demand (and still nascent preferences for higher standards for online content regulation within the CDU/SPD governing coalition) with a normative landscape that had interventionist characteristics in general, but was predisposed towards a self-regulatory strategy for online content specifically, led to a move towards collaborative platform governance via the 'Task Force.'

Because of issues of implementation and enforcement regarding these new standards negotiated via the Task Force's code of conduct, leading to a perception in the governing coalition that the platforms were not taking taking these voluntary commitments seriously, as well as the external shock of the US election and the global 'fake news' discourse, demand for change grew amongst German policymakers in 2016 and 2017. Looking back, various counterfactual policy options were on the table: Maas' Ministry could have continued to work with the existing Task Force structure, building upon the forum and framework of the code of conduct to incorporate more stringent standards or some kind of better industry auditing and monitoring, implementing some kind of data sharing or other monitoring mechanism. There were obvious ways to improve that mechanism's commitments and capabilities: for instance, the 'monitoring' that ended up being conducted by the government to measure compliance with the collaborative code of conduct was crude and unscientific, constrained by a lack of proper access to platform data and without a proper sampling strategy.[40]

Since the Task Force had gone into effect in late 2015, a few new collaborative efforts had been instigated at the European level, the most notable of which was the EU Code of Conduct on Illegal Online Hate Speech (Gorwa, 2019). While the EU Code remained largely insulated from the similar German collaborative efforts that were happening around the same time, and did not feature prominently in the domestic German debate, Maas and the executive could have joined the forum created through the Code of Conduct, trying to bring the German efforts to a broader and collaborative pan-European strategy through which to raise content moderation standards for platforms. However, collaborative platform governance was no longer seen to be good enough, given its enforcement problems and its reliance on firms to fundamentally continue calling the shots. Maas and others

[40] As German law professor Marc Leisching established through correspondence with the Ministry of Justice, the Jungendschutz team that conducted the monitoring was not composed of lawyers, and given the complex nature of some German criminal statutes, it is probable that "legal laypersons" were not actually able to identify precisely what exactly constituted illegal content under the German Criminal Code (Liesching, 2017). Likewise, the content was never archived before flagging by Jungendschutz, so it is unclear whether the flagged content was actually illegal in Germany or just removed or not removed under the platforms broader 'community standards.'

in the executive increasingly demanded change that would properly contest the status quo of private platform authority, and return foreign companies to what was perceived to be their rightful place as corporate actors bound within the democratically determined context of domestic laws and norms.

The German government was able to meet this demand and successfully supply these rules, simply, because it had the power to do so. It had the regulatory capacity to do so domestically (helped by the steady negotiation and discussion between regulators and firms via the Task Force in 2015 and 2016, but was crucially determined by the ability of the governing coalition to unequivocally command the parliament) and transnationally (exerting political pressure and will as the largest and most powerful EU state to prevent the Commission from intervening against rules which were technically against existing EU intermediary liability legislation). Finally, German policymakers saw the NetzDG as clearly within their normative remit and scope of appropriate policy intervention, drawing upon the 'Rechtsstaat' tradition in Germany legal theory (and the broader German tradition of intervening in information distribution channels, of "fighting hate with speech law," as Tworek has put it) to overcome normative concerns about potential repercussions for freedom of expression.

The NetzDG Today

The NetzDG is a vital case to examine as the world's first regulation of its type, one which shifted the global regulatory agenda for online content — with critics arguing that its approach has provided inspiration for at least a dozen other jurisdictions, including less democratic ones seeking to better control avenues for domestic dissent (Tworek and Leerssen, 2019). The successful passage of the NetzDG into German law, and its continued survival, is quite the feat when one considers that it was both a world-first effort to impose specific rules for how major transnational platform governed the political speech and activity of their users, and an effort to create new rules that clashed with an overarching legal framework in a 'regulatory state' that

not only derived power from maintaining regulatory harmonization but also saw itself as a leader on digital policy issues (Bradford, 2020; Majone, 1999).

It significantly impacted the content regulation landscape in the EU and tilted it towards increasing fragmentation. In 2019, a law very similar to the NetzDG was passed in France, although it would be overturned by the French Constitutional Court in 2020. An almost identical law to the NetzDG was passed in Austria in early 2021. In both cases, the European Commission wrote strongly worded comments questioning the compatibility of these laws with existing EU regulations, however, it did not exert its veto authority.[41] In the Austrian case, the Austrian government made explicit the argument that it would be hypocritical for the EU Commission to act in their case, given that it had failed to intervene in the NetzDG a few years earlier. But the biggest change has been that the new Von Der Leyen Commission announced the launch of a new Digital Services Act (DSA) in 2020, the first comprehensive reform of status quo intermediary liability regulations for online platforms in Europe since the E-Commerce Directive of 2000. In planning documents for the Digital Services Act and its sister competition oriented regulation, the Digital Markets Act, the Commission explicitly cites the fragmentation to the EU's harmful content-related landscape due to the NetzDG and its Member State copycats as part of the core rationale for the DSA, which would re-level the playing field and bring back harmonization (European Commission, 2020, p. 4).

Platform governance modes are also not static. Entrenching regulatory change usually takes time, and in the meantime, there can be concerted balancing act between the demand for those rules, regulatory capacity, institutional factors, and various normative characteristics that might be seeking a return to the previous status quo. The NetzDG, once in force, could have been repealed if the CDU/SPD lost the elections in September 2017, but both parties still received the most votes. The Free Democrats (FDP), one of Germany's major parties which had not crossed the minimum threshold to enter the Bundestag in 2013, were back in parliament

[41] See Gorwa FOIA to DG Grow, 2020: Commission Comments re: Austrian Communication Platforms Act. Documents received January 29, 2021; document dated December 2, 2020. https://www.asktheeu.org/en/request/commission_comments_re_austrian

following the 2017 elections. Jimmy Schulz, a newly elected FDP Member of the Bundestag, tweeted that the end of the NetzDG was coming, as the party made NetzDG repeal one of its asks for entering a coalition government (Reuter, 2017). However, by January, the SPD's talks with the FDP and Greens had ended up unsuccessful and the SPD party conference voted narrowly to join another grand coalition with the CDU, and thus the NetzDG's days were in fact not numbered at all.

In the 2017-2021 parliamentary session, NetzDG critics sought to build demand for changes to the NetzDG, with the Greens introducing a set of amendments into the Bundestag that would not go anywhere due to their lack of parliamentary influence (Rusch, 2018). The FDP also sought to appeal the law via legal channels, with two Members of the Bundestag filing a suit against the NetzDG's constitutionality in the Cologne Administrative Court, with the hope that it would be clearly struck down or referred to the Constitutional Court in Karlsruhe. However, the lawsuit was dismissed on the grounds that they were not sufficiently 'interested parties' to bring such a lawsuit; in effect, in the German system only the companies that were the regulatory targets could levy a legal complaint (Der Tagesspiegel, 2018; LVZ, 2019). This institutional structure for regulatory referral stands in a notable contrast to the French system, for example, where the Constitutional Court directly accepted a referral from one of the opposition parties that had been against the Loi Avia as soon as it was passed into law (J. Schulz, 2020).

Since the NetzDG went into effect, the assessment of its impact has been mixed. On one hand, most academic observers seem to agree that the fears presented by the law's biggest critics in civil society and industry — that the law would lead companies to massively 'overblock' legitimate and legal content on their platforms, as the law incentivized them to play it safe and remove content rather than face potential fines (Tworek and Leerssen, 2019) — have not materialized. Only one fine has been issued thus far, and firm transparency reports appear to suggest that they are rejecting a good amount of the requests they receive under the new NetzDG reporting procedures they created for users in Germany, not just blindly removing content (Heldt, 2019). However, firms varied significantly in their implementation

of the new rules, with Google, for instance, taking the spirit of the new rules and implementing them in a way that would be very user friendly for users (the NetzDG specifies a number of criminal code offenses that users should be able to flag content against; YouTube's reporting form helpfully clusters these by subject area in a user-friendly way).[42] In contrast, Facebook buried their reporting form in a number of hard-to-navigate drop down menus, in effect refusing to fully implement the measures the German government mandated (Wagner et al., 2020). Also, when requests have been made by users to flag content under German law and the NetzDG, rather than funnelling that content into a separate set of evaluative policies and practices, as intended in the NetzDG, Facebook first handles it via their usual, global process of Community Standards evaluation; only then, if it is not removed there, it will be forwarded to a smaller, separate pipeline for German legal requests (Heldt, 2019). In effect, the German effort to layer new platform governance practices on top of Facebook's has simply been subsumed into Facebook's existing processes, set up in a way that it has the smallest possible impact to Facebook's global operations. Nevertheless, it is widely acknowledged that the NetzDG led companies to increase their investment in content moderation infrastructure in Germany, hiring more local employees, increasing the sophistication of their training, and in the case of YouTube, streamlining and improving their complaints-handling infrastructure.[43]

Heiko Maas moved on from his post as Minister of Justice following the election, becoming the Foreign Minister. Nevertheless, his successors, especially Minister Christine Lambrecht, have pushed forward his broad agenda, seeking to obtain a set of amendments to the NetzDG. The first set of these has been a relatively minor set of amendments, proscribing that the complaints handling infrastructures of companies need to be 'easy-to-use' and visible (Heldt, 2020), a measure designed to remedy the perceived enforcement loophole deployed by Facebook, which buried its NetzDG complaints form. The second is a broader, more aggressive package of

[42] See `https://transparencyreport.google.com/netzdg/youtube`

[43] At the very least, this was the argument made by the interviewed platform company employees. But overall, the interviewed policymakers also acknowledged the improvements to the pre-2015 status quo that had happened in 2015-2017, whether or not they believed that this was a sufficient degree of policy change.

anti-extremist legislation that seeks once again to obtain some of the provisions in the original NetzDG draft that were removed at the behest of the Commission, including the archive of illegal content removed by platforms under the NetzDG for federal criminal prosecutors, and the removal of 'identical' or similar content via some sort of technical hashing system (see Chapter 6 for more information on how these work). The domestic demand for contested platform governance has remained high, catalyzed by the ongoing spread of far-right extremism and events like the assassination of local politician Walter Lübcke and racist anti-Semitic and anti-immigrant terrorist attacks in Halle and Hanau. At the time of writing, the first of these NetzDG amendments was passed (in May 2021); the official Bundestag press release celebrated that the NetzDG approach of contested platform governance "has fundamentally proven its worth and is to be retained" as it helps ensure that there is "No legal vacuum on the Internet."[44] The fate of the second set of amendments is uncertain, with an election incoming and major liability reforms at the European level starting to be negotiated via the DSA.

[44]See the press release available at `https://www.bundesregierung.de/breg-de/suche/bekaempfung-hasskriminalitaet-1738150`, n.p.

5

New Zealand: Collaboration and the Christchurch Call

Contents

5.1 Introduction

On 15 March 2019, a man stepped into a car full of weaponry and drove more than 300 kilometers from the city of Dunedin, New Zealand, to the larger city of Christchurch. Around 1:30pm, he posted a Facebook status with links to a personal manifesto that he had uploaded on multiple file-sharing websites, put on a ballistic helmet that had been mounted with camera designed for outdoor extreme sports,

and began a Facebook live stream, walking into the Al Noor Mosque armed with a shotgun and assault rifle (Royal Commission of Inquiry, 2020, p. 40-41). This live stream, initially seen only by a few hundred Facebook users, was quickly reported to Facebook and taken down — but not before copies were made and re-posted on internet messaging boards. Within hours, hundreds of thousands of versions of the video (some altered with watermarks or other modifications) were being re-uploaded to Facebook, as well as to YouTube and Twitter (Sonderby, 2019).

More than 50 lives were lost in the deadliest terrorist event in New Zealand's history, and the Christchurch attack provided an immediate and powerful shock to the global policy conversation around terrorism, white supremacy, and its intersections with platforms like Facebook and YouTube. Two weeks after the event, an op-ed with Mark Zuckerberg's name on the byline was published in the *Wall Street Journal*, calling for more government guidance on how platform companies should deal with harmful content. A week after that, Australian lawmakers introduced the Sharing of Abhorrent Violent Material Act, an amendment to the Australian criminal code that reformed the liability regime for online intermediaries, which now could be heavily fined (and have their employees potentially imprisoned) for failing "to ensure the expeditious removal of abhorrent violent material" online (Douek, 2020, p.4).

At the same time as the Australians were developing their approach, Prime Minister Jacinda Ardern's Labour government was trying to decide what responses New Zealand should take. While it considered a more security-focused, contested approach to shaping platform content standards like the Australians (who would craft their approach in a similar vein as the German NetzDG), New Zealand ended up instead developing a collaborative international strategy, working together with the French government to steer high-level discussions with the chief executives of the major technology companies about possible voluntary commitments. The result was the Christchurch Call, a non-binding international declaration which was announced in Paris on May 15 2019 and would eventually be signed by nearly 50 countries and 8 firms. New Zealand's sustained efforts to regulate harmful content on the largest user-generated content platforms built upon existing governmental

and industry institutional networks, and would result in some creative institutional bricolage, where a small and informal partnership of major platform companies called the Global Internet Forum to Counter Terrorism (GIFCT) would be, at the behest of the New Zealand and French governments, expanded, formalized, and spun into a private regulatory body with formal charitable and non-profit status.

The Christchurch tragedy offers a sort of policy experiment and internal comparative case study, by providing a single regulatory episode where two neighbouring countries with close economic, social, cultural, and political ties sought to respond to the same event, with the same stated policy goal, with very different regulatory approaches. Why did the responses to the Christchurch incident in Australia and New Zealand lead to divergent governance strategies? There has been very little scholarship yet to look at the AVM Act and the Christchurch Call (the notable exceptions thus far are Douek, 2020; Thompson, 2019), despite the plentiful and excellent work from Australian scholars on other elements of platform-related policy in the region.[1] The goal of the following two chapters is to help remedy this gap, and provide an in-depth look into both the Australian and New Zealand cases through a platform regulation lens.

The argument of the chapter is in effect that, although rule-makers in New Zealand had high levels of demand for regulatory intervention fueled by the crisis, and also likely had the power resources to intervene domestically, they were constrained by a normative environment that helped frame contested platform governance as a normatively inappropriate and ineffectual strategy. Due to the political culture of the governing Labour government, and the understanding of the importance of human rights and free expression held by key civil servants in positions of policy entrepreneurship, the government foregrounded human rights and civil liberties — especially the protection of legitimate forms of free expression — as a key element

[1]See for instance the work of Flew and Wilding (2020) and colleagues on the newspaper industry's clashes with platform companies and the ACCC competition policy inquiry (Flew et al., 2021), the foundational analyses of Australian intermediary liability from Pappalardo and Suzor (2018), and the helpful work on the encryption bill that preceded the AVM act (Mann, Daly, and Molnar, 2020).

of their policy response, leading them to pursue a collaborative approach even when the crisis would have likely made a hard-rule and security-focused response possible.

The chapter shows how New Zealand orchestrated a complex diplomatic effort, managing expectations across a host of different state, firm, and civil society actors, and used constant engagement — building upon their playbook from international diplomacy, and making the interesting choice to treat platform companies effectively in the same way they would other states in a multi-lateral negotiation — to oversee the implementation of new, collaboratively determined rules for how platforms governed terrorist and violent extremist content in New Zealand and beyond. Rather than contesting private platform authority, New Zealand in effect embraced it, seeking to build consensus around possible avenues of change through transnational fora with a host of different partners.

The chapter proceeds in two parts. The first outlines the regulatory context — the actors at play, their power resources and the institutional context, and the normative landscape for technology and speech regulation in New Zealand. The following section provides an analysis of the Christchurch Call's development and implementation through the lens of the conceptual argument advanced in Chapter 2.

5.2 Regulatory Context

5.2.1 Actors & Preferences

There are three broad groups of actors that played important roles in the development of New Zealand's digitally-focused Christchurch response. These were: the executive of New Zealand's elected Labour government, and a core group set up to spearhead the policy process across the Department of the Prime Minister and Cabinet (DPMC) and Ministry of Foreign Affairs and Trade (MFAT); civil society, especially the NGO InternetNZ as well as a host of other global civil society organizations; and the platform firms, which included Facebook, Google, and Twitter, as well as Microsoft.

The executive, rule-making branch of the New Zealand government is formed following elections which have historically led to governments led by one of the two major mainstream political parties, the centre-right National Party and the

centre-left Labour party. A little over two months after the German Bundestag voted through the NetzDG, and in the same week in September 2017 that voters in Germany were going to the polls to re-elect the Merkel-led CDU/SPD governing coalition, voters in New Zealand split the majority of the vote between National (which earned 56 seats) and Labour (46). Interestingly and unusually, one of the 'minor parties' of New Zealand's politics, New Zealand First, a party commonly described as populist and nationalist that had broken off from National in the 1990s (Denemark and Bowler, 2002), received enough seats (9) to make it a potential kingmaker in the post-election scrum. It decided to form a coalition with Labour, propelling Labour to a Minority government additionally backed by the Green Party (8 seats). Labour's Jacinda Ardern, elected head of the Labour party only eight weeks before the election, became Prime Minister, with New Zealand First's Winston Peters the Deputy Prime Minister; New Zealand First received four seats in the 20-member Cabinet with the Greens granted three other ministerial positions outside of Cabinet. As Vowles and Curtin (2020, p. 3) describe, this surprising turn of events "was the first time in the history of the mixed member proportional system [implemented in New Zealand in 1996] that a party with the second-most votes gained the position of leading a government."

While this elected branch sets the broad policy direction, New Zealand has various regulatory agencies and government ministries that provide input into the development, implementation, monitoring, and enforcement of rules. The central one tapped to lead the internationally collaborative response post-Christchurch was the Ministry of Foreign Affairs and Trade; other relevant ministries involved in digital policy issues included the Department of Internal Affairs (DIA), which had a team working on issues relating to online safety, the Office of Film and Literature Classification (a government agency that reviews movies, video games, and some online content), and the Ministry of Business, Innovation, and Employment, which has a mandate over telecommunications policy.

A host of domestic and international civil society groups all also played a role in the New Zealand government's response to the Christchurch Attack. While there

are few digitally-oriented civil society groups in New Zealand, the largest and most influential, InternetNZ, has a dual role as an advocacy organization as well as the technical domain name registrar for the .nz top level domain. Other groups involved in the debate around the New Zealand government's decision whether — and how — to shape platform authority included a host of the most prominent transnational civil society groups active on issues of platform governance and content moderation. Some of these groups focus specifically on digital policy (such as Access Now and the Electronic Frontier Foundation) while others, such as ARTICLE 19, work more broadly on freedom of expression and human rights issues.

Finally, the same platform actors as in the previous German chapter were the key players in the regulatory episode that follows. One notable addition to the group of Facebook, Twitter, and Google, was Microsoft, which has its global headquarters in Washington state rather than in California. Facebook, Google, and Microsoft all operate limited liability subsidiary companies registered in New Zealand;[2] Twitter only registered a local subsidiary in 2021 (Dunkley, 2021).

For the purpose of this case study, the same general (somewhat reductionist, albeit heuristically helpful) first-order assumptions about broad actor preferences can be made. The elected executive branch seeks to work towards future re-election, and the accomplishment of their stated policy goals; civil servants seek to supply the best possible rules that they can develop to meet the executive's demand for new rules, when it arises; and firms generally seek to maximize profits and maintain the lowest possible regulatory burden. Civil society groups can feature a complex set of preferences and motivations; they in many cases prefer lower standards for firms, when they perceive proposed rules to be normatively problematic; when rules fit their ideational preferences, they may argue against industry and lobby for more stringent standards through public advocacy campaigns, expert input, and other strategies (Maréchal, 2015).

[2]See the registries at `https://opencorporates.com/companies/nz/3043328` and `https://opencorporates.com/companies/nz/1786635`

5.2.2 Power Resources & Institutional Constraints

In market-sized based accounts of power, New Zealand would not be considered a powerful actor in international regulatory politics. The country has a population of about 5 million, putting it roughly around the 120th most populous country in the world, comparable in population size to Ireland, Costa Rica, Liberia, and Palestine.[3] It has a larger economy than most countries of its size, however, with a nominal GDP that would put it at about the 50th largest economy in the world. It has a very high degree of internet penetration, with 94% of residents said to be 'active internet users' in 2021,[4] as well as a extraordinarily high degree of user-generated platform penetration: according to 2018 statistics from Statista, 76% of the country's population used YouTube, 75% used Facebook, and about 30% used Instagram.[5] Microsoft's Bing search has about 3% of the New Zealand search market, which is dominated by Google Search.[6] To put these numbers in a global perspective, New Zealand's Facebook user-base in 2019 corresponded to less than one tenth of one percent of the company's global user total.[7]

Despite its small size, New Zealand has fairly strong levels of regulatory capacity. According to the World Bank's 'Worldwide Governance Indicators' dataset, which has various measures seeking to estimate "the capacity of a government to effectively formulate and implement sound policies" (Kaufmann, Kraay, and Mastruzzi, 2010, p. 4) New Zealand ranks in the world's 99th percentile in terms of its 'regulatory quality.' According to the same indices, which feature an estimate of 'government effectiveness' which measures "perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies" (Kaufmann, Kraay, and Mastruzzi,

[3] See estimates available at `https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population`

[4] See `https://www.statista.com/statistics/680688/new-zealand-internet-penetration/`

[5] See `https://www.statista.com/statistics/681512/new-zealand-facebook-users-by-age/`

[6] See `https://gs.statcounter.com/search-engine-market-share/all/new-zealand`

[7] Calculation based upon roughly 3.75 million FB users in NZ in 2019 (Statista figures) and 1.8 billion global FB users in 2019

2010, p. 4), New Zealand ranks in the global 96th percentile in 2019.[8] While these are imperfect heuristic estimates and thus must be taken with a grain of salt, it is evident that New Zealand globally is seen to have high regulatory capacity, with highly competent, well-resourced civil servants.

New Zealand, however, does not have a traditionally construed media or internet regulator; instead it has an Office of Film and Literature Classification, an independent regulator headed by a 'Chief Censor' and a small staff of fewer than 20 civil servants charged with reviewing, parental rating, and potentially barring content from New Zealand (such as films, television shows, and video games; it can also make it illegal to possess certain forms of online content, such as the terrorist manifestos or depictions of terrorist violence; Graham-McLay, 2020). As well, in 2019 the New Zealand government had a small office of about a dozen people within the Department of Internal Affairs charged with unearthing and combating child sexual exploitation material accessible in New Zealand (Kenny, 2019). This lack of large and discrete regulatory authorities with general competencies in media and tech policy (Germany has 14 media state-level authorities with some experience dealing with online content, for example), likely helped shape the decision to tap the Ministry of Foreign Affairs and Trade to end up leading the Christchurch response in partnership with the Prime Minister's Department. MFAT has about 1800 employees, more than double that of the German Ministry of Justice and Consumer Protection in charge of the NetzDG process (760 employees), although the German BMJV has a considerably larger budget than the MFAT (around 900 million Euros annually, to the MFAT's approximately 600 million Euros). MFAT does not appear to ordinarily work on technology policy issues, although it would be involved in negotiating the digitally-relevant (e.g. intellectual property) provisions of international trade agreements, for example.[9]

Institutionally, there are few formal transnational institutional constraints shaping the ability of New Zealand to intervene and supply new rules for platforms

[8]Data available at `https://databank.worldbank.org/reports.aspx?source=worldwide-governance-indicators`

[9]See `https://www.mfat.govt.nz/en/about-us/who-we-are/treaties/`

operating in the country. New Zealand is not part of a regional regulatory bloc like Germany; it also does not have any trade agreements or bilateral or multilateral agreements with the United States that place limits on the type of rules the government can deploy domestically for internet intermediaries.

5.2.3 Normative Landscape

What is the attitude to regulation in New Zealand more generally, and relating to channels of information distribution and dissemination more specifically? What is the normative landscape shaping the willingness of policymakers to intervene with rules that might have an impact on free expression?

In the mid-1980s, New Zealand began a rapid process of economic liberalization, deregulation, and commercialization (Easton, 1997). Political scientists writing in the 1990s observed the growing 'rolling back' of traditional state functions, and a move towards more minimal government intervention in the market, to the extent that the country became a "free market laboratory" for experimentation with neoliberal governance concepts (Kelsey, 1993, p. 65). In terms of media policy, as Thompson (2011, p. 11) notes, "The programme of neoliberal macroeconomic reforms in New Zealand from the mid 1980s through the 1990s saw the emergence of one of the most heavily commercialised and deregulated broadcasting sectors in the OECD," as a BBC-esque model of state-led public broadcasting was slowly dismantled. This intentionally 'light-handed' regulatory approach was also deployed in the telecommunications sector, which was 'partially de-regulated' in 1987 with the passage of the Telecommunications Act (Blanchard, 1994). Overall, even in the early 2010s, the country was still being described as having a 'laissez-faire' regulatory culture across a wide range of policy issues, especially those relating to digital policy (Barrett and Strongman, 2012), even as there was an increase in state intervention in the media and telecommunications sectors and a partial roll-back of the almost completely hands-off model of the 1990s (Hansen and Jones, 2017).

As mentioned previously, New Zealand does not have a traditionally construed media or internet regulator. It does have a long history of limited oversight in

various media and content via the — pragmatically named — Chief Censor, an independent and government-appointed official that leads on the implementation of the various film and media censorship laws that have existed in the country since 1916 (D. Anderson, 2017). The Chief Censor can append parental warnings and create new categories of 'viewer discretion advised' interstitials for films, movies, video games, and 'computer files', and has the power to render certain forms of 'objectionable' content illegal to own or share publicly.[10] However, the Censor's office is small and its true power in the online environment is limited; it was largely set-up for professionally produced content, for example for screening content to be shown at cinemas in the country. While the Censor can make online content illegal (such as a specific video, file, or piece of content, creating criminal repercussions for New Zealand citizens that download or upload this content), it did not under the legal framework in place in 2019 have the power to issue takedown notices to platform companies hosting content in other jurisdictions.[11]

Another set of newer regulations passed in 2015 also sought to tackle the issue of cyberbulling and online harassment by creating a complaints handling body to which users can submit complaints (Panzic, 2015), with the body, NetSafe, "intended to deal with complaints in the first instance" (Síthigh, 2020, p. 18) and helping intermediate between users and the platform companies that host or facilitate content. This Harmful Digital Communication Act (HCDA) does not properly constitute platform regulation, however, as as the offenses target ordinary New Zealanders and not platform companies; the new criminal offenses apply to individuals that "post a digital communication with the intention that it cause harm to a victim" (Post, 2017, p. 211), and not to firms.

While the combination of New Zealand's Chief Censor and the HCDA may make it seem as if New Zealand has a generally interventionist set of speech norms, the government has been more comfortable potentially intervening in areas such as pornography, indecency, and harassment than it is around the broader and

[10] See the documentation available at `https://www.classificationoffice.govt.nz/about-nz-classification/classification-and-the-internet/`

[11] Ibid.

more contentious areas of political speech. The New Zealand Bill of Rights Act of 1990 states that "everyone has the right to freedom of expression, including the freedom to seek, receive, and impart information and opinions of any kind in any form," without significant qualifications for dangerous and dehumanizing speech (Elers and Jayan, 2020). After a number of racist incidents in the early 2000s, Prime Minister Helen Clark's Labour government considered a set of hate speech laws in 2004, but dropped the project after political pressure and a public consultation in which the "overwhelming majority of submissions were opposed to implementing a new, general hate speech law" (Harrison, 2006, p. 71). Critics of New Zealand's laissez-faire norm around political speech note that it harms minoritized groups, underplaying the fact that "the right to free speech may negatively impinge upon the right of targeted communities to be free from discrimination, including discriminating hate speech" and that this is "a dehumansing by-product of Whiteness, racism and coloniality reflected in New Zealand's freedom of expression laws" (Elers and Jayan, 2020, p. 240).

5.3 The Development of the Christchurch Call, March - May 2019

5.3.1 Setting the Agenda and Determining Demand

In the immediate aftermath of the deadliest mass shooting event in New Zealand's recent history, policymakers in the Prime Minister's department realized that the attack had been 'designed to go viral,' and that the attacker had very cannily exploited Facebook and other user-generated content platforms in order to amplify his extremist views. This made it, quite possibly, the world's first 'internet mediated' mass shooting, and Jacinda Ardern's government needed to quickly figure out how it could respond. The conversation turned to the question of what the appropriate policy response to the online dimension of the incident should be, and multiple government departments, from the Department of Internal Affairs (which had a 'digital safety' team working on issues that included the spread of child exploitation content) to the Ministry of Justice were involved in this discussion, noting that

there were a number of potential legislative and policy gaps that had been exposed by the incident.[12] The initial tone from the government was similar to the kind of language that had come out of Germany in the lead up to the NetzDG. In her Ministerial statement to Parliament a few days after the shooting, Ardern emphasized two policy areas that had been revealed as insufficiently developed by the Christchurch Attack: gun control, and social media platform regulation. Ardern took a quite assertive tone, suggesting that the government might need to contest the way in which firms created and enforced their private regimes of content moderation: "We cannot simply sit back and accept that these platforms just exist and that what is said on them is not the responsibility of the place where they are published. They are the publisher, not just the postman. There cannot be a case of all profit, no responsibility."[13]

Nevertheless, and in contrast to the German NetzDG example, Ardern's government did not have a policy entrepreneur actively demanding new rules for user-generated content platforms even before the event. Ardern and her advisers were open to getting consultation from other branches of government, and a special team in the Prime Minister's Office working on cybersecurity and digital policy sought to provide advice to a Prime Minister looking for the best path forward. The head of this team was Paul Ash, a career diplomat with experience in cybersecurity policy, who would be one of the most important individuals in the development and implementation of New Zealand's Christchurch response (Shepherd, 2019).

This window of consultation created an opportunity for other actors to try and affect the level and character of demand for new rules. As Ash described in an interview, the executive had been "struck by the levels of engagement from industry immediately."[14] Given the crisis, and the apparent unambiguity of the content involved (which already violated the content guidelines of all the major platform

[12]Interview held via video conference with Michael Woodside, Policy Director, Department of Internal Affairs, October 2020.

[13]See the published parliamentary transcript at: `https://www.parliament.nz/en/pb/hansard-debates/rhr/combined/HansDeb_20190319_20190319_08`

[14]Interview held via videoconference with Paul Ash, Director, New Zealand National Cyber Policy Office, December 2020.

companies), Ash noted that "everyone wanted this content to come down," including industry and civil society, making interests in this policy space more aligned than similar discussions that might happen on other topics like hate speech (or areas where country laws clash with platform terms of service, as in the NetzDG case).[15]

Facebook had quickly and publicly described how they were actively trying to remove every instance of the video they could find, deploying all sorts of complex technical tools to do so (Sonderby, 2019). Microsoft's President Brad Smith also got directly in touch with Prime Minister Ardern to note his willingness to cooperate and to present a variety of policy solutions, and travelled to New Zealand immediately following the attack.[16] Smith would be a crucial policy entrepreneur in the post-Christchurch policy discussion, despite his firm not being a major player in the user-generated content space (Microsoft does not operate a social network like Google or Facebook do; and while it has a search engine, Bing, its market share in Oceania is comparatively very small).[17] Smith's conversations with the New Zealand team planted key ideas and served as a connection point between networks of government and industry higher-ups.

Digitally oriented civil society in New Zealand also set out to influence this important debate about the kinds of rules that would be demanded. As Ellen Strickland, the policy director for InternetNZ, the influential New Zealand based digital-rights NGO, described humorously in an interview, they had seen "the response from Australia and the people we know there so our first response was try and talk to our contacts in NZ governments and try and make sure that they didn't do something stupid too."[18] It helped that Jordan Carter, InternetNZ's director, had personal connections to Prime Minister Ardern through shared time

[15]Interview with Paul Ash, December 2020.

[16]See Sachdeva (2019).

[17]Microsoft's business model is more enterprise-oriented and business facing than the other firms engaged in this discussion, but it, and Brad Smith specifically, have deployed this type of policy advocacy strategically in since 2016 onward as part of an apparent effort to portray their company as more responsible than its competitors. Smith is very active on information security issues and had previously called for a 'Digital Geneva Convention' that resulted in a similar Paris-based declaration to the Christchurch Call, the 'Paris Call for Trust and Security in Cyberspace.' See Gorwa and Peez (2020) for more.

[18]Interview held via Signal with Ellen Strickland, Chief Policy Advisor (International), InternetNZ, February 2021.

spent in the NZ Labour Party (and even had once been housemates with Ardern), and thus was able to get an important early seat at the table in a way that most digital NGOs in other countries do not.[19]

Carter explained why his organization was arguing for a collaborative, non-statuatory regulatory approach in a blog post that grappled with the lack of obvious solutions to the problems exposed by Christchurch:

> Making random quick laws on our own might respond to a deep seated feeling many of us will be having that "something has to be done and NOW." The quick action on gun laws taken in New Zealand could be seen as an example on this front. Sadly, that won't work in this situation. There are no global precedents for how to deal with social media and violent extremist or terrorist content. If it was already sorted, the experience we had with Christchurch would not have happened. While it might sound painful, the right place to start is the conversation.

5.3.2 Negotiating the Options

As firm and civil society voices sought to help influence the government's demand for rules, Ash and his team eventually provided the advice to the Prime Minister that the government would need to "look at some form of collaborative and voluntary solution, and we were going to have to work internationally, including with the major tech platforms, as a way of trying to grapple with this problem and come up with constructive solutions" (Shepherd, 2019, n.p.). While there had been a discussion of prospective domestic responses, and the possibility of pursuing a more combative, contested governance response, Ash suggested in an interview that there were two main reasons as to why the collaborative strategy was chosen. Firstly, the normative implications of pursuing more interventionist online speech policies were inherently problematic for the executive. They worried about their international and domestic reputation: as Ash put it, "our country has long been saying that a free, secure, and open internet is the goal, and we couldn't do an about face on that."[20]

[19]Interview held via videoconference with Jordan Carter, Chief Executive, InternetNZ, December 2020.

[20]Interview with Paul Ash, December 2020.

The government therefore had a distinct type of laissez-faire norm shaping its self-perception of the appropriate role of intervention, seeing government overreach in this space as both possible and potentially detrimental for human rights on one hand, and also potentially hypocritical, posing prospective reputational costs on the other.

Secondly, the New Zealand government appeared to be pragmatic about its capacity to really obtain meaningful change via a contested strategy. While it is likely that they would have had the domestic power resources to pursue a domestic contested approach,[21] the executive was concerned about the costs of potentially implementing new rules and the difficulties that the small New Zealand market would have with enforcement. As Paul Ash would later put it in a media interview describing New Zealand's strategy, "We could go down the Teddy Roosevelt line, and speak softly and carry a big stick... It's just that we don't have a big stick" (Shepherd, 2019, n.p.). Alongside the normative costs of a contested approach, the New Zealanders worried that the contested strategy wouldn't work — that firms would simply exit (or threaten to exit, creating a high-profile domestic showdown) that would be extra costly for the Ardern government politically. In effect, for a combination of normative and capacity related reasons, New Zealand's executive felt that they could achieve their preferences, at least satisfactorily, if sub-optimally, by working directly with firms in a collaborative platform governance mode.

5.3.3 Developing the Text: Platform Diplomacy

On the 25th of March, a work programme for a collaborative governance initiative called the "Christchurch Call" was set by the Prime Minister's Department, with the Ministry of Foreign Affairs as lead on implementation. Prime Minister Ardern had discussed the idea for the initiative with a number of world leaders — Theresa

[21] In 2019, the Labour government would have needed support from its coalition partner, NZ First, to pass new rules. NZ First leader Winston Peters frequently re-iterated in the post-Christchurch conversation that "Prime Minister speaks for the coalition Government" (and that the party did not wish to advocate separate policy positions, especially on the sensitive post-Christchurch issues. While the counterfactual is not possible here, and I was unfortunately unable to interview any NZ First policymakers to ascertain their position on the Call or its alternatives, it appears likely that they would have supported the government had Ardern and the executive pursued a contested response. See `https://www.parliament.nz/en/pb/hansard-debates/rhr/combined/HansDeb_20190319_20190319_08`

May, Justin Trudeau, Angela Merkel, and Emmanuel Macron — and Germany and France had pointed Ardern towards the 'Tech For Good Summit' that was being hosted in Paris in May 2019 as a possible venue for an international Christchurch meeting.[22] Ash and his team travelled around Europe collecting feedback and ideas, and obtained official partnership of the French government, which not only boosted the capacity of the comparatively small New Zealand team, but also made sense diplomatically due to the location and positioning of the Paris summit which had been suggested as the international venue.[23] By mid-April, Ash's team had developed a draft of the call, that was in solid enough form to be circulated to the other stakeholders as an exposure draft.

In late April, Ash and his team travelled to California, and the draft went through its first proper negotiation with firms at meetings held in the San Francisco Bay Area. The public policy heads of the major five companies, including Nick Clegg of Facebook, Kent Walker of Google, Brad Smith of Microsoft, and David Zapolski of Amazon, along with the respective company General Counsels, attended the meeting to discuss the details. From there, the exposure draft then went through multiple rounds of revision in the two weeks before it went to Paris. This was, as Ash described, an almost 24/7 operation made possible by handing off to staff stationed in Europe at the end of the New Zealand workday. They received input from multiple corners, including from firms on what was technically feasible from their perspective; international lawyers who gave advice on potential customary law effects of the final Call, as well as its synergy with the UN Guiding Principles on Business and Human Rights ('The Ruggie Principles'). Other governments, which were being courted as potential signatories, including the United States, provided comments as well.[24]

The text that was almost fully finalized in early May was a short document, running at around 1300 words, 2-3 pages in length.[25] It was structured in four

[22]Interview with Paul Ash, December 2020.
[23]Ibid.
[24]Ibid.
[25]FOIA to the New Zealand Ministry of Foreign Affairs and Trade, August 2020. Documents recieved December 10 2020. Available at `https://fyi.org.nz/request/13466`, henceforth NZ FOIA 2020.

sections: an introductory pre-amble setting the stage ('the pledge'), followed by three sections with commitments that actors would implement to make the pledge happen. It begins with the measures government signatories would publicly agree to (e.g. "strengthening the resilience and inclusiveness of our societies to enable them to resist terrorist and violent extremist ideologies;" "Ensur[ing] effective enforcement of applicable laws that prohibit the production or dissemination of terrorist and violent extremist content"), and then moves to the commitments that firms would implement ("Take transparent, specific measures seeking to prevent the upload of terrorist and violent extremist content and to prevent its dissemination on social media and similar content-sharing services")[26]. The document concludes with a final section with joint measures that both governments and firms would publicly commit themselves to. An annotated version of the text obtained via a freedom of information request features a table breaking down the text almost line by line, discussing key negotiations and additions, as well as making plain the reasoning behind specific sections and particular wording.

The document begins with explanatory annotations that demonstrate the care that went in to crafting the document so that it would feel like a genuinely multistakeholder approach — for example, in a significant contrast to the approach being taken by the ostensibly collaborative Australian voluntary Task Force, the Call begins with commitments being made by government signatories, and not firms. As the annotations note, "We want tech companies to know this is a genuinely cooperative effort - we need to recognise the role governments play in addressing the drivers of violent extremism."[27] Additionally, the annotations explain the document's mention of government-led regulatory proposals, acknowledge that government appropriate action in this policy arena might also involve

> a full range of regulatory-type measures, from voluntary frameworks through to black letter law. This framing means that we still acknowledge the importance of domestic regulation (as countries have and will,

[26]Ibid.

[27]Annotated copy of the Christchurch Call circulated to potential supporters. NZ FOIA 2020, p. 12.

> regulate on this issue as well as alternative measures that could be designed in a collaborative way.[28]

The annotations for other parties note the Call's bounded scope ('the Call is not seeking to address all of the ills of the internet'), carefully diplomatic language about acknowledging existing work being done under the Countering Violent Extremism through existing structures like the EU Internet Forum, and outline New Zealand's "intended approach, which is to be collaborative."[29] This was a whirlwind diplomatic process which involved managing multiple relationships, ensuring that the corporate and governmental signatories would stand by the text. The document was finalized just as the New Zealand delegation flew to Paris from Sydney airport.[30]

On May 15, exactly two months after the Christchurch shooting, the Christchurch Call Summit was held in Paris. The summit included heads of government from "New Zealand, France, Jordan, Senegal, Norway, Canada, the UK, and the Vice President of Indonesia," as well as the President of the European Commission (J. Carter, 2019). The Chief Executives of major platforms Facebook, Google, and Twitter also attended, as well as representatives from Microsoft, Amazon, and Wikipedia. At the end of the meeting, the call was signed by world leaders and platform company executives. Countries supporting the Call on 15 May were New Zealand, France, Australia, Canada, Germany, Indonesia, India, Ireland, Italy, Japan, Jordan, the Netherlands, Norway, Senegal, Spain, Sweden, the United Kingdom. The European Commission also was a signatory.[31]

5.4 Implementing Collaborative Platform Governance

Unlike many other international declarations made through fora like the G20, the New Zealand government was committed to ensuring that the collaborative governance approach of the Christchurch Call would actually lead to changes in

[28] Ibid.
[29] Ibid.
[30] Interview with Paul Ash, December 2020.
[31] See `https://www.christchurchcall.com/supporters.html`

firm policies and practices. In a post-Paris debrief report, MFAT staffers summarize the four outcome areas that had been identified during the negotiation of the call text as next steps:

> Following the Paris meeting, [redacted], four priority areas for action were identified: reform of an existing industry body (the Global internet Forum to Counter Terrorism (GFlCT [sic])) to be more inclusive and effective, and take forward Call-related work; developing a shared crisis response protocol to enable countries and companies to work together better in future attacks; better understanding where there are gaps in the research on terrorist and violent extremist content online; and better understanding how companies' algorithms can drive users to more extreme content, and identifying intervention points.[32]

The document further notes that at the conclusion of the Christchurch summit, Prime Minster Ardern and French President Macron "undertook to regroup with Call supporters on the margins of UNGA [UN General Assembly] Leaders Week to assess progress against the call."[33] It was time to start the implementation process, and for New Zealand and France to assess whether the Call commitments were indeed being voluntarily and satisfactorily met. While firms made individual policy changes, the main institutional channel through which the implementation of the Call's general proscriptions would happen, however, was the Global Internet Forum to Counter Terrorism (known as GIFCT). Here, New Zealand's collaborative efforts linked up transnationally with the previous collaborative governance strategy that had been pursued in Europe by the European Commission.

The roots of the GIFCT go back to some of the earliest policy conversations in Europe about the problem of terrorist content on emerging user-generated content platforms. In December 2015, after preparatory meetings held in 2014 and 2015, the European Commission officially announced the creation of the EU Internet Forum, which brought together EU officials together with representatives from Google, Facebook, Twitter and Microsoft (Gorwa, 2019). After only six months and two meetings (that are publicly known; the entire process was deeply secretive,

[32]Pip Mclachlan and Elizabeth Thomas, 'Christchurch Call: Ministerial Progress Update.' Obtained via NZ FOIA 2020, p. 20.
[33]Ibid.

and notably excluded civil society, see Fiedler, 2015), the members of the Internet Forum announced the EU Code of Conduct on Countering Illegal Hate Speech Online, committing the firms to a wide-ranging set of principles, including the takedown of hateful speech within 24 hours under platform terms of service and the intensification of 'cooperation between themselves and other platforms and social media companies to enhance best practice sharing' (European Commission, 2016, p. 3). To comply with that commitment, the four firms announced the creation of the GIFCT in 2017 (Microsoft Corporate, 2017). The organisation, as of spring 2019, was highly secretive, and had a board made of "senior representatives from the four founding companies" and published little about its operations (Gorwa, Binns, and Katzenbach, 2020, p. 8). The goal of the informal industry best practice-sharing group was to coordinate the use and improvement of automated systems to remove extremist images, videos and text (Microsoft Corporate, 2017).

While there has been no scholarship yet to specifically examine the GIFCT and its processes, the core of the GIFCT that has been thus far been elaborated by researchers and civil society is a technical infrastructure called a hash-sharing database (Huszti-Orban, 2017; Llansó, 2016). These are systems for automatically matching content, which typically involve 'hashing,' i.e. the process of transforming a known example of a piece of content into a 'hash' – a string of data meant to uniquely identify the underlying content. Hashes are useful because they are easy to compute, and typically smaller in size than the underlying content, so it is easy to compare any given hash against a large table of existing hashes to see if it matches any of them. This is typically used to match images and video, but Facebook's policy leadership stated in 2018 that it had also begun adding text and audio to the hash-database (Bikert and Fishman, 2018), meaning that, in an impressive technical feat, effectively every type of content uploaded to Facebook — be it an image, 'status update,' or video, would be at the point of upload hashed and checked against large databases to see whether it should be blocked at upload.

Through the GIFCT database, the core firms (as well as a subsidiary group of about ten different companies that have also joined the GIFCT) can upload

and share hashes of content that they consider to be prohibited extremist material, allowing the other firms to also automatically block that content if they choose. This shared database was the core 'product' of GIFCT, but the organization also served as a talking shop, with a closed annual meeting in San Francisco that brought together firm representatives with key counter-terrorism and national security officials from the EU, US, and the Five Eyes countries.[34]

Microsoft's policy czar Brad Smith was probably the first high-level firm representative to bring up GIFCT as a potential governance instrument to move forward with post-Christchurch. In a post published on Microsoft's public-facing corporate blog a week after the Christchurch incident, he offered two concrete policy solutions: the leveraging of the existing GIFCT structures ("There are in fact important recent steps upon which we can build" when responding to Christchurch) to foster more technical capacity and collaboration within industry, and the creation of a 'crisis response protocol' that the firms could enact with government in case such a similar event were to happen again (B. Smith, 2019, n.p).

This appears to have become an increasingly popular way to implement the commitments that firms were making under the call. As one firm representative put it in an interview, "The Christchurch Call requires people to work together. And you need to have a way to do that (and build structures to do that). But rather than create something new, its easier to build up something you already got."[35] Continued discussions between the companies and the New Zealand MFAT kept the pressure on as far as implementation went. Government and firms began to negotiate the details of specific organizational changes to GIFCT that would have an impact on its core functions, policies, and legitimacy. By September 2019, the New Zealand team had been pushing hard to be able to announce some deliverables in the four months that had passed since the Paris Summit. While collaborating with the firms to oversee the overhaul of GIFCT, they were also continuing their engagement with 'foundational supporters' (those that supported the Call on 15 May), and working with France to deliver a 'second wave' of new

[34]See `https://gifct.org/about/story/#august-2017---first-meeting-of-gifct`

[35]Interview held via videoconference with anonymous platform policy manager, February 2021.

country supporters, to be announced and profiled on 23 September at the UN Meeting.[36] These included "Denmark, Mexico, Sri Lanka and South Korea, as well as UNESCO and the Council of Europe - bringing the total numbers to 48 countries and three international organisations" (McCulloch, 2019).

A Christchurch Call 'Leaders Dialogue' event was held on September 23 at the UN General Assembly, hosted by New Zealand, France, and Jordan (which had been long engaging in policy conversations relevant to terrorism through its Aqaba Process meetings). Government leaders (including Prime Minister Ardern) as well as corporate representatives (including Twitter CEO Jack Dorsey) delivered speeches, announcing three major updates: the future restructuring of GIFCT, so that it would include an 'independent advisory committee,' and would have permanent staff and a formalized institutional structure; the creation of a crisis response protocol that would be tested through a simulation event held in New Zealand in December 2019; and the broadening of the Call's membership through both a new wave of government signatories as well as a formalized 'Christchurch Call advisory network' of civil society groups.[37]

In her speech to the UN General Assembly the next day, Ardern was optimistic about these developments:

> Yesterday, I met with Call supporters to check on our collective progress. We announced that a key tech industry institution will be reshaped to give effect to those commitments – and we launched a crisis response protocol to respond to such events in the future. Neither New Zealand nor any other country could make these changes on their own. The tech companies couldn't either. We are succeeding because we are working together, and for that unprecedented and powerful act of unity New Zealand says thank you.[38]

[36]Ministerial Progress Update on the Christchurch Call, cont. Obtained via NZ FOIA 2020, p. 25-26.

[37]More than forty prominent global civil society organizations were part of this network, and as part of that, Civil society the were able to attend this meeting and offered some opportunity to speak (York, 2019).

However, as Access Now's Policy Director Javier Pallero described in an interview (videoconference, January 2021), their capacity and resources to actively contribute to the Network varies significantly. Civil society has been themselves organizing formal principles and rules for members of the advisory network, and the New Zealand government has been in conversations to equip them with a secretariat that could help offset some of the capacity burden they are facing.

[38]The full text of the speech is available at `https://www.newsroom.co.nz/full-text-pms-speech-to-the-united-nations`

As New Zealand's Paul Ash described in an interview, the work did not stop there, as the NZ MFAT team continued to work to informally monitor and pressure firms that they realize these commitments. As he noted in late 2020, it "took much longer to rebuild GIFCT than anticipated, and lots of work to make sure that they were properly staffed and equipped." This involved time consuming classic administrative work: "getting 501(c) [a legal status for a non-profit organization in the United States] took time, drafting a GIFCT charter took time, recruitment took time, and all required lawyers to be closely involved."[39]

Nevertheless, in June 2020, the newly-stood-up independent GIFCT announced an independent advisory committee, with representatives from seven governments, the EU, the UN Counter Terrorism Executive Directorate, and twelve civil society organizations.[40] It has in 2020-2021 published its first transparency report, turned its closed annual summit into a "multi-stakeholder forum," and started inviting researchers and civil society groups into working groups developing specific aspects of the GIFCT's work.[41] It remains the central (newly institutionalized) institution through which New Zealand has pursued its governance goals, and through which it seeks to shape private authority wielded by platforms not just in Oceania, but globally.

5.5 Conclusion

The goal of this chapter has been to provide a policy-oriented analysis of the development of the Christchurch Call, the collaborative platform governance initiative through which New Zealand has sought to shape the policies and practices of platform companies operating in New Zealand on matters relating to live-streaming, terrorism, and violence. To summarize, the argument presented is that the New Zealand government, motivated by normative constraints and considerations, opted to pursue

[39] Interview with Paul Ash, December 2020.

[40] See `https://gifct.org/about/story/#june-2020`

[41] Despite these changes, the GIFCT is still frequently critiqued by researchers and civil society groups. Douek (2020) has called it a 'content cartel,' and the Centre for Democracy and Technology in Washington has coordinated multiple civil society statements expressing their concerns with the GIFCT re-organization and mandate (Llansó, 2020).

a collaborative governance strategy rather than a contested one. This contrasts with the dominant 'power driven' explanation that would be the assumption of a more Dreznerian conceptual framework, and the possible argument that high enough levels of domestic demand within a state will simply yield new rules.

The overall aim of the narrative presented here is to provide an exposition of the normative 'laissez-faire' logics that I have hypothesized play a slippery, but important role in transnational platform regulation. Of course, alternative explanations may also be possible; and readers might think that the simpler, more power-based argument would be simply to say that New Zealand, due to being a small market, simply could not supply binding rules (and policymakers knew this), leading them to pursue a voluntary approach. One might argue that in effect, the norms of a free and open internet mattered less than market power. The counterargument would be that small states have in other instances been successful, to varying extents, at contesting private platform authority. For example, Singapore has a very similar population to New Zealand, but a far more assertive and interventionist (and arguably authoritarian) tendency when it comes to political speech and media and communications sectors, all of which played a significant role when Singapore developed its Protection from Online Falsehoods and Manipulation Act in 2019. Following my conceptual framework, I would argue that POFMA can be understood an instance of high government demand (albeit low market power), few institutional constraints, and an interventionist (rather than laissez-faire, as in the New Zealand context) norm. The normative terrain in New Zealand (and its position on human rights and digital issues) made a collaborative approach far more likely than in the Singapore context, despite their similar size.

Other, hard to pin down factors undoubtedly influenced the development of the Christchurch Call. As Ash noted in an interview, electoral demand factors came to play less of a role in the Christchurch Call context context, as it "was not an election year, and the government did not feel the same need to 'do something big immediately.' "[42] The calculus of normative and economic costs may have been

[42]Interview with Paul Ash, December 2020.

different for the New Zealand executive if indeed a possible election had been on the horizon. Nevertheless, many existing conversations about platform regulation are either excessively technocratic, or purely power-based. By considering other, softer, admittedly difficult to empirically measure factors like political culture, logics of appropriate action, and ideological notions around free expression and human rights, one complicates the prevailing understanding of platform regulation as merely a mechanistic response to external policy shocks.

The jury remains out on the long-term effects of the Christchurch Call as a governance instrument and institution. As Thompson (2019, p. 99) has put it, the Christchurch Call could have a major impact by representing "a very initial step toward the formation of a multilateral regulatory framework for controlling online terrorist and extremist content, along with other practices of social media and online intermediary operators." It will be fascinating to follow the effects of the Call, and the GIFCT, as well as the role that New Zealand maintains as the orchestrator of a transnational initiative that now features many larger and potentially more powerful state actors. Scholarship on public-private transnational institutions has noted that "informal governance structures may also be a source of power in their own right and may even empower otherwise weak players, such as small states and NGOs" (Westerwinter, Abbott, and Biersteker, 2021, p. 14). Will this be the case for New Zealand, or will bigger players step in to take over the institutional structures that the New Zealanders initially negotiated to better suit their own aims and goals? While collaborative platform governance has largely to date been the purview of powerful states and the European Commission (Gorwa, 2019), events like Christchurch have increased both the interest and legitimacy of non-European and small states in transnational efforts to shape platform authority, and certainly provide an important space to watch for the future of multi-stakeholder platform governance.

6

Australia: Contestation via the Abhorrent Violent Material Act

Contents

6.1 Introduction

Just over two weeks after the Christchurch Attack, a new bill titled 'The Sharing of Abhorrent Violent Material Act' was introduced into the upper house of the Australian Parliament. It was the 3rd of April 2019, and it was the last day that the Australian Senate would be sitting before breaking for a new election. Less than 24 hours later, the bill had been signed into law and the Criminal Code had

been amended include a new type of content (called "abhorrent violent material" or AVM),[1] and a host of aggressive sanctions related to the non-removal of this material by internet platforms (Douek, 2020). The penalties stipulated for failure to remove could be up to 10 million Australian dollars or 10 percent of a company's annual global turnover; as well, as Australian Attorney-General Christian Porter repeatedly emphasized in statements to the media, potential prison time for the executives of social media companies connected to this failure to remove AVM content.

In the lead up to the passage of the law, which was rapidly passed not only without consultation with the affected stakeholders, but also in the face of loud protests from industry, academic experts, and global civil society (Kaye and Aoláin, 2019; Krahulcova and Solomon, 2019), Prime Minister Scott Morrison stressed that the security-focused response was a 'world first' effort to do the hard work necessary to keep Australians (and implicitly, Kiwis and others in the region that might be again harmed by Australians like the Christchurch shooter) safe online (Prime Minister of Australia, 2019). The government publicly stressed what it saw as the appropriate hierarchy between the transnational 'community standards' and private rules created by companies like Facebook, Twitter, and YouTube and the democratically determined rules of Australia, with one government white-paper stressing that Australia needed to contest private platform authority, and "needed to regulate — to intervene in this market — for one simple reason: digital platforms have failed to act in a manner that approximates Australia's community standards."[2]

Why did the Australian government seek to pursue changes to the platform regulation status quo via a contested governance approach, when its neighbours across the Tasman Sea — who arguably had a clearer argument for a security-focused, hard-line contested approach, given that they had just experienced a terrorist attack of significant magnitude — instead opted for the collaborative

[1] As Douek (2020, p. 3) outlines, this material is that which depicts a "terrorist act, murder, attempted murder, torture, rape, or kidnapping... however, content only meets the definition where the material is recorded or streamed by the perpetrator or their accomplice"

[2] Gorwa FOIA to the Australian eSafety Commissioner and Attorney General's Department, June 2020, Document 30. Documents received September 2020, and available at `https://www.righttoknow.org.au/request/abhorrent_violent_material_act#comment-29021`. I have adopted the numbering provided by the AG's office in this chapter.

strategy? The argument advanced in this chapter is that the Australian executive, following the Christchurch shooting, demanded new rules as part of an effort to appear strong on terrorism and national security just before an upcoming election. The government was able to supply those rules due to its domestic power resources (a strong majority in parliament, and an ability to navigate the parliamentary calendar to force the legislation through in a very short period of time) and few transnational institutional barriers (including the lack of opposition from the United States, which could have used existing trade agreements to exert some pressure against the law if it wished), all enabled by a lack of constraints facilitated by Australia's interventionist normative environment around digital policy and free expression.

Following in the footsteps of the previous two chapters, this exploration begins with the regulatory context: a quick overview of the relevant actors, the power resources and institutional constraints at play, and a short summary of the relevant normative landscape. The chapter then turns to the central analysis of the development of the AVM Act.

6.2 Regulatory Context

6.2.1 Actors & Preferences

There are three groups of actors that played important roles in the development of the AVM Act. These consisted of two substate groupings within the Australian government (the executive branch of the elected Liberal/National coalition government, and the various regulatory agencies and ministries that play a role during the policy development process) and the platform companies.

The executive branch of the Australian government is formed following elections that, in the past 50 years, have led to governments being formed either by the centre-left Labor party, or by a coalition of the two major parties on the right, the Liberal party and the National party. In 2016, the Liberal/National coalition narrowly won the election, with Malcolm Turnbull elected as Prime Minister with a one-seat majority. After an internal leadership challenge, Scott Morrison became Prime Minister as Turnbull stepped down in 2018. In the lead up to the Christchurch

Attack, the governing coalition carried a narrow majority in the 150 seat House of Representatives, with Labor holding 69 seats, and a host of minor parties and independents holding the remaining 5 seats. Federal elections in Australia are held every three years, and an election was slated for mid-2019. Although Morrison had been consistently behind in the polling, which steadily predicted a Labor victory (Gauja, Sawer, and Simms, 2020), one can assume that he was heavily motivated by a desire to be elected Prime Minister and stay on as Liberal party leader.

The elected party forms the cabinet. Cabinet ministers serve as the head of the various ministries and departments that play an important role in the development, implementation, monitoring, and enforcement of rules. The most relevant of these actors in the development of the Australian Christchurch response were the Attorney General's Department, headed at the time by Attorney General (AG) Christian Porter. The AG's Department often leads on legislative drafting and is broadly concerned with matters of law and justice, but also has responsibilities over criminal law/law enforcement, and national security. Other relevant ministries included the Department of Communications and the Arts, which had policy responsibilities over the digital economy, telecommunications policy, broadcasting policy, and 'content policy relating to the information economy.'[3] The Department of the Prime Minister and Cabinet, led by the Prime Minister and his direct staff, generally determines the broad thrust of policy and also has competencies relating to the coordination of digital and 'cyber policy.' One final substate actor relevant in the enforcement of the AVM Act is the eSafety Commissioner, an independent regulatory agency that was created in 2015 to work on digital child safety issues. Their remit was expanded in 2017, and they have some competencies relating to online content, including the fight against the proliferation of child sexual abuse material online.[4]

Finally, the main actors involved on the industry side were the same cast of American transnational platform companies: Facebook, Google, and Twitter.

[3]See the Administrative Arrangements order of 2015, which outlines the scope and functions of the various departments: `https://www.legislation.gov.au/Details/C2015Q00006`

[4]See the documentation at `https://www.esafety.gov.au/about-us/who-we-are/our-legislative-functions`

Additionally, there are a few industry associations that represent the interests of the platform companies and other internet and telecommunication companies in Australia, the most prominent of which are Communications Alliance and the Digital Industry Group Inc (DIGI). Another set of industry actors that have been identified as potentially important actors in the recent literature on Australian platform regulation are the 'traditional' media companies: the broadcasting, television, and newspaper firms that compete with platform companies over advertising dollars as well as audiences. Flew et al. (2021, p. 2) have identified the potential impact that "inter-capitalist competition" between large media conglomerates like Rupert Murdoch's News Corp Australia on one hand and Facebook and Google on the other have had on some aspects of Australian efforts to regulate the platform economy, although in my analysis of the AVM Act these media firms did not appear to play a meaningful role.

6.2.2 Power Resources & Institutional Constraints

In market-size based accounts of state power, Australia would be considered a moderately powerful actor in global regulatory politics. The country has a population of approximately 25 million, making it the around the 50th most populous in the world, or slightly less populous than Madagascar, Venezuela, Nepal, and Yemen, and slightly larger than North Korea, Cameroon, and Taiwan.[5] However, it has a large economy, with a nominal GDP that would make it about the 13th largest in the world.[6] It has a high degree of internet penetration, with a steady 89% of residents estimated to have an internet connection of some sort in between 2015-2020.[7] In 2019, around 40% of Australia's population, or around 11 million people, were estimated to be active Facebook users.[8] Estimates made by social media marketing agencies put Twitter usage at about 4.6 million people, and YouTube at 15 million Australians in 2019.[9] These kind of imperfect figures would lead us to estimate

[5] https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population

[6] See https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)

[7] See https://www.statista.com/statistics/680142/australia-internet-penetration/

[8] See https://www.statista.com/statistics/304862/number-of-facebook-users-in-australia/

[9] See https://www.fiber.com.au/post/social-media-statistics-worldwide-australia

that, for the global user-generated content platform companies, the Australian market is small in the global context, although potentially more valuable from an advertising perspective given advertising revenue per user. Australia represents about 0.014% of Twitter's global user base, 0.008% of YouTube's global user base, and 0.006% of Facebook's global user base.[10]

Australia is a G20 country known to have a capable and competent bureaucracy. According to the World Bank's 'Worldwide Governance Indicators' dataset, which has various measures seeking to estimate "the capacity of a government to effectively formulate and implement sound policies" (Kaufmann, Kraay, and Mastruzzi, 2010, p. 4) Australia ranked in the world's 98th percentile in 2019 in terms of its 'regulatory quality.' According to the same indices, which feature an estimate of 'government effectiveness' which measure "perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies" (Kaufmann, Kraay, and Mastruzzi, 2010, p. 4), Australia ranked in the global 92nd percentile in 2019.[11] Australia is perceived to have high regulatory capacity, with highly competent, well-resourced civil servants, at least when considered in the global context. The Department of Communications and the Arts had 550 staff in 2018-2019, and one of the regulatory agencies under its purview, the Australian Communications and Media Authority, a converged regulator with an influence in broadcast, television, and internet policy, had an additional 427 staff (Commonwealth of Australia, 2018).

The government in Australia's Westminster-style system also has significant capacities to set domestic legislation and pursue-wide ranging policy reforms. Power is fairly concentrated in the executive: even though the Liberal/National governing coalition only had a one-seat majority after the 2016 election, it still had a huge degree of executive power, with an ability to basically pass legislation at will as long

[10]Calculation based upon 317 monthly active Twitter users in 2019; 2 billion monthly logged in YouTube users in 2019, and 1.8 monthly active users for Facebook in 2019.

[11]Data available at `https://databank.worldbank.org/reports.aspx?source=worldwide-governance-indicators`

it is able to maintain control of party members (Kumarasingham, 2013). Legislative studies scholars, discussing the similar systems of Canada and Australia, have quipped that these countries represent 'elected dictatorships' with an unusually high degree of centralization in government, not seen in systems like the United States where parties often control only part of a bicameral legislature (Sayers and Banfield, 2013). This means that in Australia, there are few domestic institutional constraints in place to bind the hands of a majority government that strongly demands new policies.

Transnationally, the only major institutional constraint on Australia's ability to supply new platform-related rules was a preferential trade agreement with the United States, the Australia-United States Free Trade Agreement (AUSFTA), signed in 2004. The agreement does contain some provisions pertaining to telecommunications policy and the digital economy, including stipulations that the signatories do not impose customs duties on digital products and do not discriminate against the 'digital products' of their trading partners (Given, 2004). This is an older trade agreement that slightly pre-dated the rise of major American user-generated content businesses (Burrell and Weatherall, 2008); it does not have more specific considerations directly mandating the types of intermediary liability provisions that the signatories should implement or maintain, as seen in the 'NAFTA 2.0' US-Mexico-Canada Trade Agreement of 2018 (Krishnamurthy and Fjeld, 2020). AUSFTA thus provided a relatively weak constraint that the Australian government could likely overcome as long as they were able to make a coherent argument that any new rules they supplied did not merely discriminate against American firms.

6.2.3 Normative Landscape

What is the attitude to regulation in Australia more generally, and relating to channels of information distribution and dissemination more specifically? What is the normative landscape shaping the willingness of policymakers to intervene with rules that might have an impact on free expression?

Scholars of Australian regulation have used various conceptual tools to explain the evolution of its regulatory philosophy in the past 50 years, but it appears as if Australia has historically embodied a more interventionist 'regulatory state' approach to regulating firms and markets than many other G20 countries. Rather than total neoliberalization and de-regulation, the 1980s and 1990s were associated with (often large-scale) reforms in key economic sectors such as finance, leading to a certain degree of "regulatory liberalisation and privatisation" (Allen et al., 2021, p. 118). In the media and broadcasting sectors, there has always been significant state involvement: As B. Goldsmith and Thomas (2012, p. 2) note, it has been observed that "the history of Australian content in broadcasting has been a history of regulation," and there is a long tradition of Australian governments intervening in information distribution channels for social and cultural reasons.

Part of the story is likely that, as the Australian media regulation scholar Terry Flew has pointed out, the absence of constitutionally guaranteed rights in Australia has meant that the balance of limits to rights (including to expression and speech) versus security has been frequently tilted towards security, especially by conservative governments (see also Mann, Daly, and Molnar, 2020).[12] This has included an interest in online safety, including attempts to set limits on the types of online content that Australian residents are able to access, and the things that they can say and do via platform services. In 2011, a report by the Australian Law Reform Commission, an independent regulatory authority that conducts periodic assessments into existing Australian legal frameworks, recommended new legislation that required internet intermediaries to "block or remove 'prohibited' content available on or through their networks" (Pappalardo and Suzor, 2018, p. 470). In 2015, the Tony Abbott-led Liberal/National government implemented the Enhancing Online Safety Act, which created the office of the eSafety Commissioner (eSafety), which advertises itself as "the world's first government agency committed to keeping

[12]Interview held via videoconference with Terry Flew, Professor of Digital Communication and Culture at the University of Sydney, December 2020.

its citizens safer online."[13] eSafety's mandate was initially confined to child safety, and issues like cyberbullying and child abuse material, but in 2017, Prime Minister Malcolm Turnbull's Liberal/National government expanded its remit to include "online safety for all Australians."[14] In the lead up to the Christchurch attack, eSafety had a reporting mechanism for online content, and had the power to issue takedown requests to companies in certain cases.

This is all to say that one can argue that policy debates in Australia are underpinned by a relatively interventionist normative frame about the appropriate scale of government intervention in the (digital) public sphere. While this landscape is constantly evolving and being contested — there are also political actors within the country seeking to move towards a more libertarian, US style model of free expression, that have sought to secure the repeal of the most recent iteration of Australian hate speech legislation passed by Labor in 1995 (Gelber and McNamara, 2013) — the last three Liberal/National governments have been very muscular on security policies and terrorism, especially on its intersections with the digital, and have been willing to sacrifice both privacy and free expression as a result. In 2015, the Turnbull Liberal/National government enacted a set of serious telecommunications rules requiring that service providers retained metadata (information about calls, who called whom, etc) about their customers for two years for possible counter-terrorism investigations (Sarre, 2017). Relatedly, as Mann, Daly, and Molnar (2020) describe, Prime Minister Turnbull quipped that in June 2017 that "the laws of mathematics are very commendable, but the only law that applies in Australia is the law of Australia," setting off a broad national debate about end-to-end encryption that culminated in significant legislative reforms seeking to increase the ability of security agencies to conduct counter-terrorism and intelligence operations by intercepting communications on Australian networks. These comments about the unequivocal primacy of Australian law over all else (including perhaps the natural

[13] See `http://web.archive.org/web/20210302074810/https://www.esafety.gov.au/about-us`

[14] See `http://web.archive.org/web/20210304070612/https://www.esafety.gov.au/about-us/who-we-are/our-legislative-functions`

world) presaged the very similar arguments made by other Australian political leaders around platform companies and their content moderation practices in mid-2019.

6.3 The Development of the AVM Act, March-April 2019

6.3.1 Weighing Early Options and Setting the Agenda

When news began to spread that the Christchurch shooter was an Australian citizen, the Australian government needed to decide how it would respond. As Prime Minister Morrison issued a short statement on the day of the attack, the rest of his Department had to weigh potential options. The eSafety Commissioner's Office, which had experience with seeking to prevent deeply problematic and illegal content like child abuse imagery, prepared legal advice for the Prime Minister's Department with a summary of the existing Australian legal frameworks and how they applied to the attack. The advice, which was discussed by officials via email, summarized the challenge as follows:

> Australia has a robust domestic Classification Scheme for films, computer games and certain publications. [...] Under the Online Content Scheme, the eSafety Office can take action in relation to material hosted in Australia that has been assessed against the National Classification Scheme as 'prohibited' or 'potential prohibited' [...] The RC category includes offensive depictions or descriptions of children and illegal content. However, it is important to note that what is considered prohibited/potential prohibited under Australian law may not be illegal in the jurisdiction where the content is hosted.
>
> While the eSafety Office does not have the power under the Online Content Scheme to issue a takedown notice to Facebook, which is based in the United States, it does work cooperatively with digital platforms to request removal of material that is clearly illegal in Australia and other jurisdictions.[15]

As officials in the Communications ministry made edits to this advice from eSafety and prepared policy recommendations and 'next steps' for the Prime Minister's office, they noted a number of challenges facing any policy seeking to

[15] Gorwa FOIA to the Australian Attorney General's Department, Document 1.

govern how companies made content moderation decisions. Firstly, the jurisdictional tensions: because Facebook and most other content hosts were located outside of Australia, they believed that issuing direct takedown orders would have limited effect — "Domestic regulation can only go so far in addressing this as digital platforms are global entities."[16] Secondly, other targeted efforts to prevent similar types of content from spreading again would be difficult to implement: "Prohibiting live streaming is not feasible as this functionality is widely available across any number of social media, OTT and telecommunication platforms."[17] The summary of the legal context also noted that the major companies (Facebook and Google) had been working rapidly in an effort to take down the shooter's video, and appeared to be doing their best to remove this content that was against their formal Community Standards as well as their commercial and political interest.

Nevertheless, the Department of the Prime Minister and Cabinet, when responding to journalists over the weekend following the attack, took a more assertive tone, starting to set the agenda for domestic command-and-control regulation. Morrison's statement began,

> There has been a sea change in the attitude of the community and governments to the regulation of the internet over the last decade. The clear view of our Government and the Australian community is that the same standards and rules that apply in the physical world should apply in the online world. The internet cannot be an ungoverned and safe space for terrorists and other criminals.

After mentioning the track record of Liberal-National's tough-on-digital-issues agenda ("The Australian Government has been at the forefront of online safety legislative reform to enshrine the principle that the online world is not a safe place for terrorists. It's why we have legislated to give law enforcement agencies [...] crime fighting tools for encrypted communication [...] established and appointed the world's first eSafety commissioner [...] and legislated the world's first kids' anti cyberbullying regime to give the eSafety Commissioner the powers to issue take down notices and fine individuals and digital platforms"), the remarkable

[16]Ibid.
[17]Ibid.

statement noted that more needed to be done, and that the government would "not hesitate to legislate as it has in areas such as encryption, kids' cyberbullying and the non-consensual sharing of intimate images" (Prime Minister of Australia, 2019, n.p).

In the statement, the Prime Minister noted that he would be calling a meeting on March 26th 2019 to discuss these issues directly with representatives of the technology sector to discuss ways forward. At this meeting, various possibilities for an Australian policy response would be discussed. The statement concluded on an assertive note: "A best endeavours approach is no longer good enough. It's clear that while social media companies have cooperated with authorities to remove some of that disgusting content, more needs to be done. If they won't act, we need to."[18]

6.3.2 Early Negotiations: The Brisbane Summit

The goal of the March 26th summit, according to the official invitation, was for "Summit participants [to] work collectively to identify what can be done to prevent the streaming and reposting of extremist material, both now and into the future."[19] The briefing paper prepared before the meeting noted that the "objective of the Summit is to get a commitment from the digital platforms and telecommunications industry that they will lift their game and do more to deal proactively and decisively with inappropriate content."[20]

Industry began circulating some of their key arguments for the meeting. In an email sent on the weekend before the Brisbane summit with firms and government, a Microsoft staffer sent a pre-publication version of a blog post written by Microsoft President Brad Smith to the official coordinating the policy response within the Department of Communications and the Arts, asking for input and advice on whether the tone of the blog was appropriate.[21] In the post, Smith acknowledged the role that platform companies had played in the dissemination of the Christchurch shooter's video and manifesto, and offered some policy solutions while making a few important

[18]Gorwa FOIA to the Australian Attorney General's Department, Document 3.
[19]Ibid., Document 4.
[20]Ibid., Document 30.
[21]Ibid. Document 16.

agenda-influencing moves of his own. Smith offered up two concrete voluntary self-regulatory proposals. The first was to foster increased industry collaboration on terrorist content via a (then) little-known entity called the Global Internet Forum to Counter Terrorism (GIFCT). The second proposal was the development of a shared rapid response mechanism: "the tech sector should consider creating a 'major event' protocol, in which technology companies would work from a joint virtual command centre during a major incident."[22] Australian officials liked the post so much that they discuss including it as part of the briefing pack for the summit.[23]

Via email, an inter-agency discussion of the bargaining position of the government laid out the following changes desired in the rules affecting how firms governed their platforms:

> As discussed, can each agency turn their mind to tangible outcomes and changes we would propose to platforms and ISPs. As per the discussion, we propose that these outcomes would be grouped under the following elements:
>
> 1. Instantaneous or quicker takedown of violent and extreme material (or blocking of access); 2. Improving transparency of the actions the planforms and ISPs take in relation to violent and extreme material; 3. Holding platforms, ISPs, and individuals to account for the upload and distribution of violent and extreme material.

The Brisbane Summit was two hours long. It was attended by the key members of the cabinet (Prime Minister Scott Morisson, the Minister for Communications and the Arts Mitch Fifield, the Attorney General Christian Porter, and Home Affairs Minister Peter Dutton), and representatives from the three major user-generated content platforms (Google, Facebook, and Twitter), four of the major Australian Internet Service Provicers (e.g. Vodafone, Telstra), and a representative of Communications Alliance, the digital industry lobby group.

After an introduction from the Prime Minister and an overview of the structure and 'expectations' for the discussions with the Ministers, the platform companies were each allotted time to discuss their response to the Christchurch incident, the

[22]Ibid.
[23]Ibid.

failures of their relevant self-regulatory 'rules and standards' which were highlighted by the incident, and what lessons they had learned for the future. Government then sought to get specific commitments that firms would change their policies and practices around how they moderated harmful content: the talking points for Minister of Communication Mitch Fifield noted that

> we are seeking action in three areas: Prevention and protection – including detecting, blocking, and instantaneous and faster takedown options for violent and extreme material. Transparency – improving transparency of the actions taken by platforms and ISPs in relation to violent and extreme material. Deterrence – enhancing responsibility for the upload and distribution of violent and extreme material by individuals, platforms and ISPs. Today we are seeking concrete actions and commitments from industry.[24]

The briefing packs circulated to the participants featured specific asks from the platform companies that had clear similarities with the core asks of the German Task Force that had had similar conversations four years prior, including more detailed transparency reporting on how the platforms conducted their moderation, the acquisition of more moderators domestically within the country, and the streamlined takedown of illegal content.[25] A number of the industry representatives that attended the summit said that the outcome from the negotiations with government at the meeting was positive: that they would develop a collaborative code of conduct and set of best practices that would affect their content moderation policies and practices domestically.[26] The platform companies agreed to form a task force with officials from the Ministries to develop these informal rules voluntarily, with the additional aim of providing input into prospective additional amendments to the Enhancing Online Safety Act, the regulatory framework which had been drafted in 2015 and created the office of the eSafety Commissioner.

[24]Ibid., Document 8
[25]Ibid.
[26]Interview held via videoconference with Christiane Gillespie-Jones, Director (Program Management), Communications Alliance, December 2020.

6.3.3 Demand for Contestation Builds

The summit happened on Tuesday, March 26, 2019. It looked as if the government and the firms had agreed to pursue changes to content rules via a collaborative governance strategy. However, following the summit, email summaries of the day's discussions noted that a new major deliverable, not included in the original meeting agenda or briefing materials, was being fast tracked: a set of new platform-related criminal offenses and sanctions being developed by the Attorney General's department. The last day of parliament before dissolution for the 2019 federal election was the next Thursday, April 4th. It appeared that Attorney General Christian Porter, in consultation with Morrison and perhaps the broader cabinet, decided that they might have a crack at trying to draft a rapid new legislation to point to while electioneering.

It is difficult to pinpoint exactly when this strategy came into effect, but executive demand for a contested platform governance approach with binding rules suddenly increased. While the pre-Summit public communication discussed the possibility of the government stepping in if necessary, the post-Summit statements framed this intervention as effectively a done deal. A press release put out by the Prime Minister's Department on the 30th of March was emblematic of this newly assertive approach. Prime Minister Morrison offered a number of explications of the way that they would 'force' firms to do what they perceived to do the right thing: "Big social media companies have a responsibility to take every possible action to ensure their technology products are not exploited by murderous terrorists," Morrison said. "It should not just be a matter of just doing the right thing. It should be the law. And that is what my Government will be doing next week to force social media companies to get their act together" (Prime Minister of Australia, 2019, n.p).

Parliamentary Debate and Opposition

This new bill, titled the Sharing of Abhorrent Violent Material Act (commonly abbreviated as the AVM Act) was introduced into the upper house of the Australian Parliament on the 3rd of April, about two and half weeks after the Christchurch

attack. The day was scheduled as the last session for the Senate to sit before the election, the agenda was incredibly packed, and party politics and last-minute maneuvering and electioneering abounded. The ordinary legislative process is for a bill to be introduced, read, and passed in one of the two houses, and then read and passed in the other house in identical form. According to statistics from the Parliament of Australia, approximately 95 percent of proposed laws are first introduced in the House (site of most amendments, debate, and discussion) before going to the Senate for approval.[27] Because of the way that the parliamentary sitting calendar for 2018-2019 had been drafted, however, with a shortened schedule, it happened that the Senate had its last session on April 3rd followed by a final House session on April 4th. For this reason, the Liberal/National government was unusually — and in a turn into the realm of informal governance (Kleine, 2013) — introducing a number of bills into the Senate on the 3rd with the aim of passing them in the House the next day.

As the leader of the Australian Green Party (which at the time held 9 out of the 76 Senate seats) stated in comments a few minutes after the session began at 9:30 am,

> The Senate's been on strike for the past few months and now we're being asked to support 30 bills, ramming them through this parliament with the support of the Labor Party. We haven't even seen some of these bills! We have not even seen the bills that will be rammed through this parliament. We're dealing here with some legislation that will fundamentally change people's lives. Let's look at what we're actually being asked to support.

Senator Di Natale outlined his critiques of three new pieces of legislation that he saw as being introduced last minute before the election with inadequate consultation and discussion: a deeply problematic welfare reform bill, a fossil fuel infrastructure bill, and, as he put it,

> Then we've got some of the most significant changes to social media online regulation that we have ever seen. This bill hasn't even been introduced. It hasn't even been introduced and it's going to be rammed

[27] See the documentation at `https://www.aph.gov.au/About_Parliament/House_of_Representatives/Powers_practice_and_procedure/Practice7/HTML/Chapter10/Bills%E2%80%94the_parliamentary_process`

> through. We haven't had an opportunity to see it. Of course, in the wake of Christchurch, we need to look at how we regulate social media and online content. Of course we need to do that. People shouldn't be subjected to the abhorrent material that's posted online. But you don't go about this by introducing legislation that the parliament can't even debate and scrutinise.[28]

Twelve hours later, the session was still going, and the Liberal government was still trying to push through bills at a tremendous pace. The AVM Act was introduced at 9:13 pm, and two minutes later, the next item on the docket was already being discussed, with the bill magically having gone through a first and third reading. The official Senate record documents the slightly comedic confusion and frustration of Senators who are unable to follow which bills are being introduced and voted on.[29] As one crossbench Senator complained to the media the following day, the bill was never actually read or properly introduced. "It is bad enough when the government forces a vote on a Bill that members of the public haven't had a chance to respond to. But in this instance, even the Senators haven't had a chance to look at it" (Duckett, 2019).

In the House of Representatives the next day, Attorney General Porter gave a speech outlining the core of the new policy approach, which involved the amendment of the existing Criminal Code to include a new type of content, called "abhorrent violent" material.[30] Two criminal offences were proposed in the amendment, relating to this new type of content. The first is 'a failure to notify' the Australian Federal Police about abhorrent violent material circulating in Australia when there are "reasonable grounds" for believing the acts being depicted happened in the country.[31] The second offence relates to companies that 'fail to remove' abhorrent violent material propagating via their services 'expeditiously.' These offenses apply not only to large social media companies, but also potentially to complementor firms

[28]Senate Hansard for April 3, 2019, p. 819

[29]See `https://www.openaustralia.org.au/senate/?id=2019-04-03.225.1&s=Sharing+of+Abhorrent+Violent+Material`

[30]As Douek (2020, p. 3) outlines, this material is that which depicts a "terrorist act, murder, attempted murder, torture, rape, or kidnapping... however, content only meets the definition where the material is recorded or streamed by the perpetrator or their accomplice"

[31]See the legislation at `https://www.legislation.gov.au/Details/C2019A00038`

that provide technical or other infrastructure, as well as internet service providers and telecommunications companies (Douek, 2020). They were underpinned by significant punitive sanctions: the penalties stipulated for failure to remove could be up to 10 million Australian dollars or 10 percent of annual global turnover; as well, as Attorney-General Christian Porter repeatedly emphasized in various statements, potential prison time for executives of social media companies connected to this failure to remove. In the speech, the Attorney-General deployed securitizing language, framing social media platforms as responsible for hate, terror, and violence, and framing a 'tough on platforms' stance that was the only appropriate response to such a crisis event.[32]

In various statements, multiple members of the opposition Labor party pointed out potential flaws with the hastily-drafted legislation and complained about the lack of serious scrutiny that it had received, noting that the legislation had not even been circulated to members of the opposition. While Labor's Shadow Attorney-General critiqued the process and the policy, Labor ended up supporting the bill, knowing that its lack of a House Majority meant that it was in effect powerless to stop it.[33] The party stated that it would vote for the bill with the caveat that it would amend it after the election if elected, a strategy that they hoped would reduce the chance that Labor would be tarred as having voted against efforts to curb violent extremism and 'abhorrent violent material' during the upcoming election campaign.

The bill was voted on and passed via the Liberal/National majority on April 4th. The Prime Minister framed the legislation as a symbolic centerpiece of the Australian government's Christchurch policy response, a 'tough on Big Tech' stance that would not just provide a signal to industry, but also to voters in the lead up to an election. "It's a very strong message and we are not mucking around," Morrison was quoted as having said (Lynch, 2019).

[32]See https://www.openaustralia.org.au/debates/?id=2019-04-04.15.1&s=Sharing+of+Abhorrent+Violent+Material#g15.2

[33]See https://www.openaustralia.org.au/debates/?id=2019-04-04.15.1&s=Sharing+of+Abhorrent+Violent+Material#g15.2

6.4 Assessing and Implementing the AVM Act

The rules governing how American user-generated content platforms in Australia policed content were thus changed quite significantly in just over two weeks. The emergence of the legislation appears to be well explained by supply and demand based frameworks of change: after Christchurch, the demand for these rules to change, at least symbolically, became interlinked with the Liberal/National coalition government's pre-electoral strategy. Both the Prime Minister and the Attorney General, immediately after the Brisbane meeting on March 26th (which also signalled the exploration of possible collaborative options with industry) began to deploy securitizing rhetoric, framing their strategy as an ambitious 'world first' effort to stand up to these powerful foreign multinationals. This demand led the Attorney General's department to draft new legislation in effectively a week (even although less costly collaborative alternatives were being developed in parallel by the Communications Ministry in the shape of the voluntary 'Task Force'). The government was able to supply these changes and get them through parliament due to domestic power resources, in particular the Liberal/National control of both houses of parliament. Other regulatory capacities played a role as well, including the staffing and capabilities to draft new legislation in an extraordinarily short period of time (six working days), and to work the legislative system to the extent that the bills were able to be so rapidly introduced, voted upon, and advanced. In effect, the governing coalition bullied the new Criminal Code Amendments through the legislature; they were helped by a lack of domestic institutional features that would have made this more difficult, such as legislative choke points for the opposition to 'filibuster' or stall, or any other kind of time and process constraints.

The remarkable aspect of the development is that the Australian government was able to develop and legislate such a potentially sweeping regulatory change in such a short period time; and perhaps, some power-based accounts of corporate governance might additionally note, that they were able to do so at all. Australia is not a world power in terms of economic might or market size in the way that the US or the EU is, and one would expect major disadvantageous changes to the

platform liability regime in the country — changes underpinned with potential fines up to 10 percent of global turnover or employee imprisonment — to be strongly lobbied and fought against by industry. Additionally, given concerns that the bill might problematize US-Australian trading and security relationships under the Five Eyes umbrella, this was a potentially high cost and risky type of regulative change, on top of the economic and innovation costs that industry articulated.

The firms were especially frustrated, the clear loser of a process in which it had not had time to provide any input or really mobilize against the changes. As Sunita Bose, the managing director of the Digital Industry Group, an industry-interest advocacy and lobby group that represents Facebook, Google, and Twitter in Australia put it,

> Let's be clear: No one wants abhorrent content on their websites, and DIGI members work to take this down as quickly as possible. But with the vast volumes of content uploaded to the Internet every second, this is a highly complex problem that requires discussion with the technology industry, legal experts, the media and civil society to get the solution right. That didn't happen this week.
>
> This creates a strict Internet intermediary liability regime that is out of step with the notice-and-takedown regimes in Europe and the United States, and is therefore bad for Internet users as it encourages companies to proactively surveil the vast volumes of user-generated content being uploaded at any given minute (Cameron, 2019).

Nevertheless, the government demanded new rules, and saw a window of opportunity to do so. There were few constraints on this demand: in terms of transnational constraints, AUSFTA was insufficient, with its lack of specific liability provisions and a perception within the Australian government that it was too outdated, and that the internet had changed so much in the 15 years since it had been drafted, to meaningfully proscribe what the Australian government could do in terms of rules for the digital economy.[34] Additionally, there were no real normative constraints to curb the scope of appropriate intervention from the Morrison government, which securitized the issue and framed it discursively as part of a broader dual-pronged strategy of (a) re-asserting Australian sovereignty over

[34]Interview held with Nic Suzor, Professor, Queensland University of Technology, November 2020.

foreign tech companies and (b) portraying the online domain as unambiguously part of the Australian government's remit. Under these conditions, the Australian government was able to achieve a strategy of contested platform governance and pass binding rules with an impact on how private platform firms policed user-generated content in the region.

6.5 Conclusion

The goal of this chapter was to provide a process-oriented analysis of the development of the AVM Act, the contested platform governance initiative through which Australia sought to shape the policies and practices of platform companies operating in Australia on matters relating to live-streaming, terrorism, and violence. The argument presented is simple: that the Australian government demanded new rules (largely due to strategic electoral and branding considerations) and was able to supply those rules, as it had the requisite power resources domestically, and few transnational institutional constraints to prevent it from doing so. Additionally, the government's interventionist normative frame made it more likely that they would pursue a contested platform governance strategy, even as collaborative options were available.

There are many ways in which this story could have potentially been a very different one: each one of the factors I have outlined as potentially of interest in my conceptual model could have been the site of greater contestation and conflict. If a major industry organization had really been able to lobby more strongly against the rules — or if major media organizations like News Corp came out against them — perhaps demand would have been tempered and the legislation would not have been introduced in the 2016-2019 session. Similarly, If the firms had been able to mobilize the US government's representatives in Australia (or better yet, the State Department in Washington) to engage in some back-channel discussion with their Australian counterparts to pressure against the new laws, or ever perhaps suggest that they might consider retaliatory sanctions of some sort (or investigate how these new rules might contravene the Australian-US trade agreement), demand may have

also been quelled. Other possible demand contestation points would have involved public advocacy and relations campaigns led by civil society or industry, but it seems unlikely that this could have happened in the matter of days in which the legislation was developed and deployed, and digital civil society groups in Australia are arguably not as prominent or powerful as their counterparts in the EU or the US. The scenario would have also been different if Australia did not have a political system which in effect prevents Members of Parliament from voting their personal opinions, rather than being 'whipped' and voting as a bloc. It would have been much more difficult for the governing coalition's one-seat majority to persist under such an institutional structure, and one can feasibly imagine at least some of the Liberal (or National) MPs peeling off or voicing reservations with the law on free expression or due process grounds.

One of the most interesting parts of this story is that, at least in part, the Morrison gamble worked. On the 18th of May 2019, Australians headed to the polls, and the Liberal/National coalition government won a surprise, 'miracle' victory, despite having consistently trailed in most major polls for the duration of its three year term (Gauja, Sawer, and Simms, 2020). This, evidently, meant that the AVM Act was not going to be referred to a parliamentary committee for further evaluation and amendment, as promised by Labor and desired by the Green Party. Instead, the AVM Act was on the books and firms needed to decide how to respond.

The speed at which the law had come into play — with the platform companies thinking that a voluntary code of conduct would be the outcome of the March 26 summit only to find out a few days later that legislation was incoming — had kept firms from being able to properly mobilize resources against the law or even suggest amendments or changes. But after the election, as one policy employee working for a major platform company jokingly put it in an interview, they quickly had a "come to Jesus moment." They continued:

> The penalties are severe — fines, imprisonment — and that started to sink in. As well, operationally, we had the question of how do we implement this: [the law] seems to suggest that we would need to do

> lots of predictive and proactive screening... does this have a proactive monitoring obligation?[35]

Interestingly, one of the main vehicles through which firms sought to respond to the AVM Act was through the voluntary Taskforce which eventually came out of the Brisbane Summit. The Taskforce's industry members, Facebook, YouTube, Amazon, Microsoft and Twitter, as well as the telecommunications providers Telstra, Vodafone, TPG and Optus, worked with the Communications Ministry and the Attorney General's Department to negotiate a number of voluntary recommendations that would help with the implementation of the AVM Act. Firms in effect were able to propose and discuss potential internal policy and process developments that would help them meet the AVM Act's proscriptions, and reduce the risk of them being selected for prosecution (and therefore potential fines) by the Attorney General's Department.

Commitments discussed in the Taskforce's report included the recommendation that firms conduct specific and bi-annual transparency reports relating to violent extremism in Australia, and ensure that clear appeals mechanisms to challenge moderation decisions were available. Additionally, the report recommended that firms develop stricter rules over live streaming, such as thresholds (e.g. numbers of subscribers) governing how one could use a platform's live stream feature, a policy change that had been publicly desired by Australian ministers.[36] In May 2019, Facebook announced that it was changing its global livestreaming policy, so that accounts which had been found in violation of other policies (for example, had content removed that promoted one the banned groups on Facebook's list of 'Dangerous Organizations') would be unable to use Facebook Live, Facebook's live video-streaming tool. Google also changed YouTube's live streaming rules globally, adding a simple floor, where channels with less than a thousand subscribers would be unable to stream live (making it far less likely that an unknown lone wolf would

[35]Interview held via videoconference with platform policy manager, January 2021.

[36]At March 26 Brisbane summit, Communications Minister Fifield reportedly asked platforms why they could not require 24 hour notice from accounts seeking to live-stream, or implement other types of friction or delay before allowing for streaming (Hunter and Duke, 2019).

be able to broadcast a mass shooting, for instance). The broader effects of the AVM Act, however, remain unclear, and as of writing, there have yet to be fines issued against the firms for non-compliance.

Following the 2019 election, a group of civil society organizations joined forces with a group of platform companies to seek possible amendments to the AVM Act. Google funded and spearheaded the campaign, which involved meetings with MPs and various government representatives; although the group wrote various recommendations and suggestions for possible amendments, these went nowhere.[37] A review of the legislation and its effects is slated for later in 2021, two years after it went into force, and may recommend changes to the Act's implementation, monitoring, or enforcement procedures.

[37]Interview Samantha Yorke, Google Australia, January 2021; interview with Nic Suzor, QUT, November 2020.

7 Conclusion

Contents

7.1 Summary of the Argument

This thesis sought to answer the following research question: *What explains the variation in how governments intervene in platform governance*?

In the preceding chapters, I have argued that government actors seek to shape how platform companies govern their services in either a collaborative or contested fashion. When demand for change is sufficiently high, governments seeking to change the status quo may in some cases deploy binding, domestic rules that attempt to shape firm activity in their jurisdiction, and in other cases, collaborate with firms, engaging in co-regulatory or otherwise voluntary efforts to incentivize firms to change their content moderation regimes to better meet state demand. I argued that the emergence of collaborative versus contested platform governance

strategies can be explained by a combination of three broad factors: sufficient demand to intervene, sufficient power to intervene (understood as domestic power resources minus any transnational constraints on the ability to supply new rules), and underlying normative considerations affecting the willingness of policymakers to intervene via binding rules.

I then sought to explore the plausibility of these arguments by looking at three case studies, selected out of a possible universe of eleven regulatory episodes identified in the mapping exercise in Chapter 3. Germany was selected as the first identified instance of contested platform governance, and thus a key case; New Zealand and Australia were additionally selected as providing internal comparisons that would help illustrate the divergence between collaborative and contested governance strategies.

	Germany	New Zealand	Australia
Demand for Change	Yes	Yes	Yes
Regulatory Capacity	Sufficient	Sufficient	Sufficient
Transnational Constraints	Yes	No	Some
Normative Environment	Interventionist	Laissez-Faire	Interventionist
Outcome	Contested	Collaborative	Contested

Table 7.1: Macro-Overview of Case-Study Features

The German and Australian cases have a number of similarities. In both episodes, the executive had high levels of demand for change, and sought to achieve that change in a relatively short period of time bounded by an election campaign. (The German NetzDG and the Australian AVM Act were both passed through parliament on the last day it sat before an election.) In the German case, the high levels of domestic demand were not tempered by concerted opposition from German and international civil society groups, or by lobbying from the companies. This high demand, combined with Germany's power resources domestically and within the EU, helped overcome the not-insignificant matter of EU resistance, and allowed Germany to supply the

	Germany	New Zealand	Australia
Substance	New obligations re: 20 existing criminal code sections	Non-binding declaration	Two new criminal offenses relating to 'AVM'
Penalties	5 million EUR	N/A	10% of annual turnover
Reporting	2 reports per year	N/A	N/A
Scope	'social networks' w. more than 2 million users	Signatories	internet, content, or hosting' service providers

Table 7.2: Overview of new rules. Penalties are upper bounds; reporting requirements are for firms in scope.

rules that it demanded — and that key policy entrepreneurs saw as normatively within their purview — despite the tensions these rules created with existing EU legal frameworks. In the Australian case, the law was deployed so quickly that it neutralized many possible efforts from firms or civil society to reduce government demand; the transnational institutional constraints were additionally far fewer, although the executive had to consider the possible ramifications for US-Australian trade discussions and the (fairly outdated) Australian-US free trade agreement which had comprehensive copyright provisions but little regarding other forms of user-generated content or modern platform services. Normatively, interventionist Australian lawmakers saw the AVM Act as clearly within their remit, and it fit comfortably within the security-over-free expression digital policy agenda of the Liberal/National government that was in power.

The New Zealand case shows how other types of outcomes are possible. Although New Zealand would have likely had the moral legitimacy following the Christchurch attack to seek a more interventionist, contested approach to governing online content, and I argue that it had the domestic demand for change and likely the power resources to do so domestically, a laissez-faire norm around the 'free and open internet,' and the importance of preserving free expression without excessive government control, shaped the New Zealand government's scope of policy action. For that reason, the New Zealand executive worked collaboratively with Google, Facebook, Microsoft, and Twitter to develop the text of the Christchurch Call, and then continued to engage actively with the firms so that these commitments would be channeled institutionally into reforms of a public-private transnational governance initiative called the Global Internet Forum to Counter Terrorism.

7.2 Theoretical Takeaways and Contributions

The underlying motivation of the conceptual framework presented in this thesis, and of the arguments presented through the case studies explored in the preceding chapters, has been to discuss the ways in which platform governance does not just have some political features (as a growing number of scholars coming from science and technology studies and digital media and communication studies have recently argued), but to go further by exploring the ways in which platform governance outcomes are actively shaped in a process of global political contestation. This thesis has sought to contribute a theoretical sensibility informed by global regulatory politics scholarship to highlight the importance of a host of domestic and transnational political factors in the interactions between large multinational technology companies and states of various sizes as they negotiate the acceptable bounds of speech, behaviour, and private power online. In other words, I have argued that to understand platform regulation, one must look beyond just the platforms and how they have historically developed their processes of content moderation across various policy issue areas, as recent work has admirably shown (Gillespie, 2018; Klonick, 2017; Suzor, 2019); one must also consider domestic political factors such as elections, and their timing; party preferences, policy entrepreneurs, and their strategies; power resources, political resolve, and broader institutional patterns and constraints. Additionally, transnational regulatory institutions and the negotiation of informal sources of regulation across different jurisdictions and sets of actors have increasingly become part of a landscape that also features a broad set of powerful and potentially unique normative features around the appropriate trade-offs between various rights and freedoms online. It is only natural that the process of platform regulation — given that it is has become a matter of public policy — should be understood as a political process, but this perspective has been largely missing from existing interdisciplinary work on the topic.

Overall, I believe that the three cases assessed here have presented adequate support for the conceptual argument presented in Chapter 2, despite a few specific instances at which it may be stretched to its limit. Explaining complex policy

processes through a few conceptual lenses offers analytical clarity but can miss the ways in which these various elements are intertwined and interconnected. For example, in the New Zealand case, to what extent did the executive's domestic political standing — which at the time was a minority government, and thus did not possess an absolute ability to easily pass any new rules that it wished, as it relied on support from its coalition partner — play a role in its eventual decision to pursue the Christchurch Call, which it then framed through a normative lens? To what extent was a contested approach ever demanded? Another thoughtful reader might note that the importance ascribed to norms in this account is primarily explored via the New Zealand case. Is it not possible that New Zealand's eventual pursuit of collaborative platform governance was not due to normative characteristics but rather to its relatively scare power resources when compared to its larger Australian neighbour? This important point demonstrates some of the potential shortcomings of the approach that I advance here, where the core factors motivating change in these case studies are qualitative and matter of argument rather than of empirically verifiable fact. Nevertheless, other existing cases outlined in Chapter 3, such as Singapore's Protection from Online Falsehoods Act (POFMA) law, show that small states with similar market power as New Zealand have been able to pursue contested strategies when they have the domestic power resources and interventionist norms to do so. In this sense, I believe that given sufficient domestic power resources, there is no reason why small states cannot pursue contested strategies; of course, whether those strategies succeed in obtaining the desired outcomes is another matter.

Overall, the original framework articulated in this thesis provides a toolbox that can guide explanations of the emergence or non-emergence of state-led platform regulation initiatives, as well as variation between them. Why have there been few successful efforts to contest platform authority in the United States, despite apparent increases in demand for the government to do so, represented not only in polling data but also in media coverage and expert debate?[1] An explanation using the conceptual lens outlined here might explore the levels and sources of demand:

[1]See for example `https://www.pewresearch.org/fact-tank/2020/12/10/fast-facts-on-americans-views-about-social-media-as-facebook-faces-legal-challenge/`

have policymakers been captured by industry or had their preferences shaped by industry lobbying; does the executive still remain largely pro-platform, despite the presence of parts of the Democratic and Republican party that seem to being demanding higher rules; and what are the power resources domestically (legislative gridlock makes it much more difficult for new rules to be supplied without bi-partisan support)? Also, what are the institutional constraints transnationally: for example, recent US administrations have put language into new trade agreements, such as the US-Mexico-Canada trade agreement, which seems to bind all parties to maintain permissive liability rules (Krishnamurthy and Fjeld, 2020). Finally, what does the normative landscape look like, and what are the issues posed by the normative First Amendment tradition that platform companies were birthed in? All of these factors might be explored in depth in future work, and key regulatory episodes in which regulation did not occur (as in the recent US context), or in which a collaborative governance strategy could have, but did not, become a contested strategy, would help further test and refine the arguments presented here.

Overall, the thesis has sought to make two specific contributions to international relations scholarship. The primary theoretical contribution was to explore a potentially novel form of private authority without formal delegation, what some developing work by Srivastava (2021) has identified as 'governance without authority' or formally granted consent. I argued here that the state-firm contestation over private platform rule-making does not fit neatly within established conceptions of either entrepreneurial or delegated private business authority that is most common in global politics, nor does it clearly fit within the existing concepts of private governance without consent like the 'illicit' authority wielded by armed groups in areas of limited statehood. The aim of the thesis was not to provide a universal theory of this type of phenomenon, nor to conceptualize and theorize how precisely this form of platform rule-making authority should be understood, in a historical context, as similar or dissimilar to the ways in which other types of influential global political private actors function. Instead, it has been to provide a first descriptive overview of the phenomenon and how it has been shaped or contested by

states in the last several years. Nevertheless, this may be a fruitful angle for future exploration by scholars of global governance interested in new forms of private authority. As certain technically complex sets of private rule-making are becoming more salient due to mass digitization and globalization, the patterns of global communication networks create risks for governments — exposing their citizens to digital content that can be harmful at a individual or collective level — these interdependencies are increasingly being challenged and contested by the states seeking to manage those risks. While the leading edge of work on globalization under these conditions of 'weaponized interdependence' (Farrell and Newman, 2019) has focused on how governments use these interconnections to exert economic and political power, the burgeoning case of platform governance demonstrates how states are also pushing back, seeking to shape private authority to better suit their ends, whether that be by asserting sovereignty and the primacy of domestic laws and norms or by championing a more collaborative approach grounded in a more inernationalist normative language of rights and freedoms.

The secondary empirical contribution has been to extend some elements of international relations and global politics literature into a hereto largely under-explored domain. By providing three new case studies, and a host of new data on regulatory initiatives both formal and informal, the thesis has brought transnational regulation work in conversation with an understanding of platform companies and content moderation that comes out of the existing interdisciplinary body of work from legal and communications scholars. The main implication has been to show that concepts from IR have purchase in the platform governance arena, and that platform companies — as companies existing within the international system, not just ethereal technology 'platforms' without physical footprint or political affect — need to navigate state efforts to assert their authority in order to continue their day-to-day operations internationally. While this thesis has only provided a relatively high-level overview of the process of political contestation between states and firms, it has sought to show that platform companies are at the very least an interesting future object of study for scholars of global regulation both

public and private. Future theses and books will need to grapple with strategies deployed by other actors not discussed in detail here (for example, platform company lobbying, which remains under-explored and under-theorized; civil society campaigns targeting platforms, which seem to be influential within the discourse and yet again have been understudied).

7.3 Project Limitations

Like any doctoral project, writing this thesis involved overcoming multiple empirical and methodological hurdles. While conducting this research, difficult decisions about scope were made at a number of points that increased the feasibility of the study but came at the cost of a certain level of generalizability and replicability. Firstly, in terms of topical scope, this work brought with it a relatively narrow understanding of online content regulation, and did not engage substantively with how multinational companies for user-generated content may be regulated in domains like intellectual property, competition policy, or data protection. Some aspects of these policy areas have content dimensions as well, and the platform companies do engage in platform governance in them — setting rules and practices around, for example, what kind of potentially copyrighted content can be uploaded to a service like YouTube or Facebook (Erickson and Kretschmer, 2018; Perel and Elkin-Koren, 2015). I believe that the basic framework of my argument still applies when it comes to copyright or the right to be forgotten (a jurisdiction demands rules, and will seek to supply them if has the power to do so) but the normative and institutional dimensions are significantly different. Following my transfer of status, I decided to limit the scope of my study to just that of 'harmful content,' a tough call which was made easier to swallow by both the prevalence of excellent IR work that focuses in on those data protection and intellectual property debates (Farrell and Newman, 2019; Haggart, 2014; Tusikov, 2016), as well as the relative degree of separation I empirically observed between the regulatory initiatives explored in this thesis and those other, broader content regulation domains. Firms largely organize themselves to meet these varying regulatory pressures along separate lines

— for example, Facebook has a 'Director of Global Content Regulation' who seeks to design the firm's strategies for responding to regulatory demands in various jurisdictions in a holistic manner; he deals with harmful content related regulatory proposals but not copyright or intellectual property.[2] The governments I observed also seem to keep these policy discussions separate, with different ministries and actors involved in the development of copyright policies than those working on what here I conceive more narrowly as platform regulation.

I also focused on a relatively narrow set of companies, mainly the most globally relevant and perhaps most politically visible platform firms — for the most part, Facebook, Google/YouTube, and Twitter — and thus did not engage much with regional platforms which have some regulatory salience in certain countries and contexts (e.g. VKontakte in Russia; WeChat in China). I also focused, out of necessity, on user-generated content platforms; I have not significantly engaged with the many politically salient platform companies with other business models, such as the operation of infrastructure (providing cloud services, for example), large commercial marketplaces, or the provision of 'gig' labour and service platforms, that are part of separate regulatory processes than those discussed here. Those platforms are also increasingly important parts of the global economy (and local economies), and have been subject to a growing academic and policy conversation (Culpepper and Thelen, 2019; Plantin and Punathambekar, 2019; Woodcock and Graham, 2020) that could also benefit from a regulatory politics oriented perspective.

The second large set of limitations that is worth mentioning relates to this thesis' methods and research design. The issues discussed in this thesis cannot be abstracted away into high-level quantitive analyses as it addresses emerging, socially and politically context dependent topics with a relatively small number of cases. In terms of research design, I had to make difficult choices about case selection, and needed to limit case studies to a number feasibly researched within a doctoral timeframe. This means that for this project, I could not undertake a fully random, large-scale sample of cases as is sometimes considered the gold standard

[2]Interview held via videoconference with Tony Close, Director of Global Content Regulation, Facebook, April 2021.

for case-study based research (Seawright and Gerring, 2008). For that reason, the regulatory episodes chosen for their salience and ability to provide within-case comparisons of institutional change over time, probably offer 'most-likely' rather than 'least-likely' cases for my theoretical model. Nevertheless, the goal of this work has been to advance a conceptual framework for this issue area that can be further expanded in future research.

Any research design largely based on interviews also comes with its own methodological limitations. An especially important one is access: in this thesis, I have only been able to include publicly available information, and information that I was able to obtain 'on the record' from interviews or from freedom of information requests. In effect, this limited my reach to information that I could triangulate with multiple publicly available and robust sources. Many of the interviews I conducted with policymakers and industry employees pertained to politically or commercially sensitive topics, and as a non-industry or government affiliated researcher, I could not expect to obtain access to the most sensitive, secretive internal deliberations around a regulatory moment. This is an unfortunate but inevitable aspect of qualitative work, and I sought to remedy it where possible by using tactical freedom of information requests (see Methods Appendix A for a deeper discussion).

Access is related to the question of sampling: who are you interviewing? I was unable to get some of the interviews with key government stakeholders that I would have liked (for instance; the German Ministry of Justice's head civil servant on NetzDG, Gerd Billen), and the narrative I have told here in the case studies was inevitably coloured by the perspectives of the interviewees with whom I spoke. Similarly, platform companies are famously cagey and usually unwilling to go on the record with journalists and researchers. They are corporate entities with a bottom line and public image to curate, and at earlier phases of this project (as seen in my transfer of status document) I initially had hoped for deeper access to the policy-making staff at Facebook than I was successfully able to obtain. Although I had hoped that I would be able to use more sensitive, micro-level interviews and perhaps even participant observation to better understand how industry policies and practices

were shaped by regulatory pressure from governments in the cases I examined, the initial access that seemed possible for my project in 2018-2019 eventually evaporated — largely because it was linked to another project I had been working on with a high-profile and senior collaborator at Oxford (Garton Ash, Gorwa, and Metaxa, 2019). Following the completion of this project, the firm expressed little interest in providing continued access for research conducted by a mere PhD student.

To side-step these challenges of access, I got creative in how I procured documents, using FOI requests to obtain deliberative documents from public institutions. However, no corresponding mechanism for acquiring policy documentation exists for companies. Partially for this reason, my thesis focused on publicly verifiable processes of regulatory change as motivated by governmental actors, and not on the internal and opaque processes of regulatory change inside platform companies, although I have sought to inject publicly available reports and press releases from firms into these cases where firms have provided some detail. Freedom of information requests provide an under-utilized tool to get new information and to triangulate interviews; however, these laws also generally include significant latitude for policymakers to redact and withhold the most sensitive documentation (Walby and Larsen, 2012). Given all these factors, it is not impossible that some components of my argument — or details of my case studies — might be contradicted by data that is classified or not public and that I was unable to obtain.

Overall, I believe that the limitations that I have faced throughout this project have been outweighed by the various conceptual and empirical contributions of my thesis. Despite the relatively narrow topical scope, the analysis presented in this thesis, and its high-level conceptualization of regulatory regimes in the digital policy context, has a number of insights that have the potential to carry over to other important neighbouring issue areas. While the specific theory here is based around this particular (important) issue area, the overall conceptual approach used for applying regulatory politics and global governance concepts should have some insights for other Internet Governance and emerging technology debates. On the methods side, despite the clear issues faced by working in a rapidly developing, often

'black-boxed' policy arena, the cases presented here still go into far more extensive analytical detail than existing work, improving our understanding of the dynamics at play in a number of key transnational regulatory moments. Additionally, the analysis of both new and existing data sources in Chapter 3 provides a macro-level snapshot of this regulatory arena, which, while necessarily cursory and limited in various ways (see Methods Appendix A for more on the limitations of the specific data and the coding strategies used), is still significantly more comprehensive than any work that has been published to date.

7.4 Avenues For Future Research

This thesis has sought broadly to provide the most complete existing examination of the transnational regulatory politics of regulating potentially harmful user-generated content. It is a fast-moving and rapidly developing topic that sits at the intersection of many disciplines, and thus offers countless potential avenues for related research going forward. A few particularly promising analytical and empirical opportunities jump out.

First, the combination of political science concepts and methods is a particularly promising one, which could yield numerous productive lines of inquiry. Given that the discussion of platform companies fundamentally pertains to companies — how they are regulated, and the varying geopolitical, economic, and jurisdictional tensions that come into play as a result — regulatory politics scholars, international political economists, and others in other political science sub-fields have much to contribute. There has been some recent work channeling this "regulatory politics approach to platform governance" (Gorwa, 2021), including the work of Medzini (2021) on platform's self-regulatory regimes and that of Haggart and C. I. Keller (2021) on the legitimacy of recent governance efforts, but their is ample other low-hanging fruit on this broad topic that has yet to be picked.

From an international relations and global governance perspective, one ambitious project would be to try and more closely link domestic changes in what Farrell and Newman (2010, p. 611) call "international market regulation" — rule changes made

by domestic actors which then have international consequences — more directly to precise changes in the rule-making and setting capacities of the large platform actors in question. What kind of research designs can help us better understand the conditions under which domestic regulatory change leads to international/transnationally translated change on behalf of firms? Here, one might consider the kind of global changes in policies and practices that can have a major impact on the lives of billions. The challenge is that platform companies are highly opaque and secretive about how they make their rules, guarding their institutional histories closely; however, the steady drip of increasing transparency on behalf of firms (Gorwa and Garton Ash, 2020) might eventually make that kind of extra-detailed process tracing possible.

Relatedly, in light of the long body of work that has sought to more closely scrutinize and assess major regulatory powers like the European Union, and situate its regulatory capacity in the global context (Bradford, 2020), more work is needed to better understand the transnational economic and political dynamics between the EU, the United States, and other countries interested in platform regulation. To what extent does the United States really go to bat for its firms, exerting pressure on countries that are seeking to make major changes to the status quo? This has yet to meaningfully happen in the case studies discussed here, but the impending EU Digital Services Act, which promises to re-wire the foundational intermediary liability rules upon which platform companies depend, provides a clear opportunity for those kinds of dynamics. To what extent will the US be content for the EU to effectively be the global rule-setter for platform rules? Will firms seek to maintain their 'one world, one rulebook' policy, where they maintain a single, relatively homogeneous set of global content policies and practices, or are their services divisible enough that different sets of rules and practices will emerge in different jurisdictions?

Additionally, the general research agenda outlined in this thesis provides an opportunity to look beyond just the domain of user-generated content. What are the dimensions of platform governance across various issue areas and types of platforms? How are the strategies deployed by the biggest, most global 'GAFAM' platforms in response to regulatory pressure similar or different those deployed by

platform companies more active politically at the municipal and local level, such as Uber or AirBnb? How do the companies operating marketplaces for goods or labour fit politically within the broader platform regulation environment? How do these firms deploy private authority, and how has their evolution followed — or not — the patterns outlined in this thesis?

A second batch of areas to explore relates to the huge amount of empirical ground which has yet to be covered. As this is such a rapidly evolving area, with a multitude of new and important government-led regulatory initiatives already on the horizon, it will be hugely important to continue to assess these emerging cases with detailed case studies. Additionally, while this thesis focuses on the interaction between states and firms, more work is clearly needed to further explore the role that civil society and other actors play in framing and seeking to influence these debates.

I personally would like to dive further into some of the cases which I was not able to get to in this thesis, including the Loi Avia in France and the POFMA in Singapore, both of which have been in effect for a sufficiently long time to allow for a detailed retrospective case study. Additionally, I would like to expand and more holistically ground my framework with more Global South cases, including looks at the debates going on in Brazil, Tanzania, South Africa, and Ethiopia. Under what conditions does neutral platform governance persist, and why has there not been as much contestation and collaboration led by these countries? A larger book project growing out of this thesis could potentially add a number of additional case studies to the three addressed here.

7.5 Looking Ahead: Policy Developments on the Horizon

The transnational regulation of online speech via platform companies is an extraordinarily vibrant and fast-moving policy area. It is an exciting, but also difficult topic to research given the constant policy innovation and change taking place. New regulatory initiatives, whether orchestrated by states or by firms, seem to be an almost monthly, if not weekly occurrence.

If this thesis was being written a decade ago, it likely would have struck a very different tone, and looked at the way in which firms were setting politically relevant rules either on their own or in occasional collaboration with various governments. Much scholarship of this period was instinctively optimistic about platforms like Facebook, YouTube, and Twitter, portraying them as socially beneficial, and perhaps inherently democratizing and liberating (Diamond, 2010; Owen, 2015; Tucker et al., 2017). A major part of this argument hinged on content moderation and platform governance: these platforms were liberating not only due to their affordances for connecting politically engaged populations around the world, but also crucially because they provided a space governed by different (privately determined, often American or American-inspired) norms and rules around expression than those that were the 'offline' status quo in many politically or socially repressive locales. In 2021, the majority of that scholarship seems, through advantages offered by hindsight, to have been almost hopelessly naive about the role that private power, and powerful corporate actors, would eventually play in public life around the world. The pendulum of optimism and non-intervention amongst wealthy and democratic Global North countries that not long ago were demanding internet freedom appears to have has strongly swung towards increased techno-skepticism and an increased interest in public, rather than private, control and oversight.

For this reason, I have sought to be as dispassionate as possible in the presentation of these case studies and in the general theoretical depiction of my argument. I have sought to not make judgements about the normative desirability of certain forms of government intervention in platform governance, but only to observe that it is increasingly taking place, and provide a framework through which one might understand why that is. In five or ten years, the pendulum might indeed swing back the other way, or new entrants (or technologies) might enable significant change that as of now cannot be fully predicted or forseen.

For now, it looks like the current trend towards increasing public contestation in the platform governance domain is only going to intensify. Two notable recent events include the election of Joseph Biden to the US presidency. Biden has in

the past made statements indicating his belief in the need to amend fundamental intermediary liability frameworks in the United States,[3] and the introduction of the Digital Services Act (DSA) package in Brussels. The DSA, which in effect constitutes a re-negotiation of the baseline intermediary liability rules which affect all online intermediaries, is in particular shaping up to be a hugely high-stakes political battle, with reports of major lobbying mobilization from the platform companies having already surfaced (Corporate Europe Observatory, 2020). Along with the Online Safety reforms in the United Kingdom, and new rules on the horizon in India, Brazil, Canada, and elsewhere, online content moderation around issues of potentially harmful speech — and not just copyright and certain types of privacy infringing content — have significantly increased in their political salience across the globe.

Other trends to note include a steadily building trickle of international coordination between smaller and medium states on the topic of platform regulation and governance. Some of this is occurring via the 'International Grand Committee' of parliamentarians from countries like Argentina, Australia, Brazil, Canada, Costa Rica, Ecuador, Estonia, Finland, France, Germany, Ireland, Latvia, Mexico, Morocco, Singapore, St. Lucia, the United Kingdom, and the United States, who have met to agree upon strategies for international cooperation and harmonization around the regulation of private platform rule-making.[4] Other state-led initiatives are being taken forward via established multilateral fora like the G20 and through organizations like the OECD. Firms are additionally experimenting with flashy new self-regulatory bodies, such as the Facebook Oversight Board, an independent legal entity set up with a initial grant of 130 million USD from Facebook, which is intended to provide Facebook with policy advice and a degree of independent adjudication over certain especially thorny takedown decisions (Klonick, 2020).

The future seems to promise a far greater, and far more existential, degree of firm-state contestation over the parameters of large platforms' global rule-making authority. But might models like the Christchurch Call become more prevalent as some states, driven by normative considerations or other types of power constraints,

[3]For example see Kelly (2020).

[4]See more details about the IGC at `https://www.cigionline.org/igc/`

seek instead to incentivize firms to develop their own initiatives like the Facebook Oversight Board or the Global Internet Forum to Counter Terrorism? Or will governments continue to deploy blunt political will and power in an effort to 'take back control' over the private standards that are increasingly having major political, social, and economic impacts on the lives of their citizens? The trend appears to be more collaboration and more contestation — in effect, more intervention of both kinds — from governments around the world. Will the outcome be fairer, more accountable, and more transparent systems of platform governance? Or will the outcome be territorialization and fragmentation as firms are pulled to splinter their services and rules to comply with a vastly increased regulatory burden across jurisdictions, perhaps with deleterious effects for vulnerable populations, political mobilization, and free expression around the globe? The stakes are high, but the consequences for both platform and state power remain unclear. Nonetheless, these dynamics of contestation and collaboration in platform governance are just beginning to unfold, and promise to become an increasingly salient, important, and impactful dimension of global governance in the years ahead.

Appendices

A

Methods Appendix

Contents

This appendix provides a more comprehensive discussion of methods than included in the introductory chapter. Part 1 features a more detailed breakdown of the qualitative interview process I underwent, as well as a discussion of ethical issues (and ethics approval via the CUREC process), a discussion of participant attribution and data use, and a list of the interviewees that agreed to have their names publicly listed in this appendix. Part 2 discusses all other types of data deployed, including the strategy for obtaining new documents via freedom of information requests, a list of archived links at which all of the obtained documents are available, and a discussion of the dataset assembled for the analysis in Chapter 3.

A.1 The Interview Process

A good portion of the original empirical material assessed in this thesis consisted of qualitative interview data, collected in semi-structured conversations with participants in the regulatory episodes scrutinized as case studies. Interview data can be potentially extremely rich as a source of new, previously under-or-un reported information, but also comes with a special set of questions, concerns, and potential pitfalls. Major questions for a researcher to tackle include: who does one talk to, how does one convince them to talk, and how is interview data handled, parsed, and attributed?

These questions are part of mainstream political science methods discussions (Mason-Bish, 2019; Morris, 2009), but are potentially even more acute when studying an area as characterized by both corporate and government secrecy as the one tackled by this thesis. Platform companies were for a long time shrouded in secrecy, operating as closed "black-boxes" despite their ostensibly public-facing nature and self-professed belief in workspace transparency (Gorwa and Garton Ash, 2020). As Maréchal and Roberts (2018, p. 2) outline, these firms "tend not to have a culture of transparency, and resist sharing what their consider to be their proprietary data," making qualitative research into company practices especially difficult. While there have been a few examples of researchers in the last few years managing to negotiate extraordinary access to companies like Facebook (see e.g. Klonick, 2020), these have been rare, almost one-off occurrences, and researchers who gain access into Facebook's operations are either bound with non-disclosure agreements or committed to staying off the record. On-the-record qualitative interviews with platform company employees remain rare, and notable recent books on content moderation that rely on industry interviews not only usually anonymize their interviews, but even may anonymize the names of the companies in question (Roberts, 2019).

I wished to also conduct interviews with policymakers and government stakeholders, as well as some key civil society organizations that had sought to influence the policy processes in question. Governments are frequently also cagey about how

they interact with platform companies, as negotiations can be political and ongoing; that said, there is a wide range of potential government interview targets (elected representatives and their staff, civil servants in various departments) and I expected current or former policymakers to be slightly more willing to be interviewed than platform employees. Civil society also provides an invaluable potential source of knowledge, often both as in-depth observers of regulatory processes in areas where there is strong digitally oriented civil society, and as creators of advocacy and journalism that documents regulatory processes. Using a broad definition of the non-governmental and non-corporate actor category like that outlined by Abbott and Snidal (2009), researchers and academic experts that participated in regulatory debates were also considered as potential interview targets. I figured that this broad set of civil society actors would be the most open to being interviewed, but also unfortunately constituted the least powerful and consequential actor group during regulatory negotiations.

A considerable literature discussing the best practices for various types of interviews exists, tackling important questions such as the kinds of sampling strategies that should be used in the social sciences, and what constitutes an adequate number of interviews (Berry, 2002; Leech, 2002). In some studies, which seek to, for example, interview a representative sample of participants in a particular universe (such as an online forum, or specific social group), an especially precise and careful sampling strategy may be needed (Eynon, Schroeder, and Fry, 2009). Because I had a clearly defined universe of participants (those involved in the regulatory episodes that constituted my case studies), my sampling strategy was a purposive one, that effectively came down to an effort to interview as many of the most relevant individuals as possible during the limited amount of time available to me during the research phase of this thesis. I benefited greatly from my personal networks and 'snowballs' pushed downhill by key interlocutors and colleagues; for this reason, I cannot of course claim that my interviews represented a complete, or representative sample of key individuals involved in the policy discussions, especially as I was unable to get access to a number of potentially important high-level

individuals; nevertheless, I did my best to identify individuals that played important roles within their respective actor group and seek interviews through them by either contacting them directly or leveraging colleague introductions.

My task was complicated by the COVID-19 pandemic. Many potential interlocutors fell ill and went on leave, or were unable to participate in an interview due to work-from-home childcare responsibilities and general business. Nevertheless, while I had originally intended to conduct face-to-face interviews where possible, I was forced to shift to a virtual interview strategy due to the stay-at-home situation, which had the major benefit of allowing me to fairly easily conduct multiple interviews with participants in different cities or even different countries in a single day. The work-from-home pandemic situation also normalized home work and videoconference calls to a certain extent, and I was able to secure video calls with some high level interviewees that, I assume, would have normally not found the time to do so. When starting each case study, I often began by interviewing relevant civil society groups and academics, asking them to also provide me with a shortlist of key individuals involved in the case study events; many were so kind as to directly introduce me to government/industry people that they knew personally, and these direct connections were generally much more likely to yield an interview than a cold email.

In the end, I conducted 52 semi-structured interviews from the end of 2019 to early 2021. Some of these were quite short, around 30 minutes, and others were much longer, closer to an hour and a half or two hours, with an average length of about an hour. The majority were conducted remotely via videoconferencing software (and in a few cases, with only audio, when the internet connection was unstable or if participants preferred to have their video off). I offered participants the opportunity to choose the platform of their choice, and conducted interviews via secure encrypted services like Jitsi and Signal as well as well as services like Zoom, Google Hangouts, Microsoft Teams, Cisco Webex, and BlueJeans.

	Germany	Australia	New Zealand	Totals
Government	13	3	2	18
Industry	3	3	4	10
Civil Society	10	7	7	24
Totals	26	13	13	52

Table A.1: General Overview of Interview Participants, by Chapter

A breakdown of the interview participants can be found in the table above. This is a broad overview, where I sought to highlight the main focus point of each interview as well as the main 'hat' that participants wore, although some interviewees fit into multiple categories, and some interviews were wide ranging and touched on multiple case studies. This was especially true in the New Zealand and Australia interviews given the interlinkages of the case.

With those caveats in mind, a few general observations about my interview process are apparent from this table: firstly, I conducted the most interviews for the Germany chapter, and was also able to get the best access to policymakers involved in the Germany case study. There were a few reasons for this: I moved to Berlin in MT 2019 to research this chapter, and held a few visiting fellowships (at the Weizenbaum Institute, Humboldt Institute for Internet and Society, and the WZB Berlin Social Science Centre) over the course of my stay in Germany that helped me build personal networks that went a long way in getting some access to good informants for interviews. Additionally, the surface for potential interviewees was comparatively larger for this chapter, as I also wished to gain insight into the negotiations between the German executive branch and the European Commission during the NetzDG discussions; this provided a large potential pool of individuals involved in these conversations from the EU side, including EU Parliamentarians and staff in the European Commission. Although I had been hoping to travel to Australia and New Zealand, the pandemic made that impossible, and relatively speaking, I had far fewer contacts and networks there to draw upon — especially in government — than I did in Europe. The relatively low number of interviews here is also a reflection of time limitations, and my need to shift into writing mode

to finish this thesis in the mandated timeline. It is definitely a limitation, and I hope to conduct some additional interviews in the future for the journal articles I hope to publish with the Australia and New Zealand chapters.

Despite my relatively strong contacts in this research space, built over four years of attending academic, policy, and industry-led conferences and workshops, I still found it difficult to get access to the appropriate industry employees. Partially this is a function of knowing who to talk to: because there is no directory of who to contact (some governments, like Australia's, have publicly available directories of government employees) or general contact points, the entire process is effectively based on personal networks and connections. Many platform employees that I contacted (including those that I had personal connections to) either did not respond or politely let me know that they were too busy in an extremely stressful pandemic time. Many others declined to be interviewed. Nevertheless, a number of high-level individuals in the companies at the global level, as well as key regional representatives, were generous enough to give me some of their time and answer some questions. Industry interviews tended to be the shortest, with some firm employees scheduling me in for a half-hour call that could be slotted in to their busy schedules.

A.1.1 Research Ethics and Attribution

Ethics are an incredibly important part of any research project, especially those that involve qualitative data and human research subjects. Following the best practices in digital research (the Association of Internet Researchers Guidelines 3.0) and political science research (the American Political Science Association Ethics Guidelines), I did my best to ensure that my interview subjects and the data collected from them were handled with the appropriate level of care.

One classic ethical issue involves the potential re-identification of participants. It is traditional best practice in qualitative research to ensure that research-subjects are properly anonymized, given the possible harms and embarrassment that can occur as a result from published material (Markham, Buchanan, and Committee, 2012), and that clear processes and guidelines for attribution are followed. This was

understandably especially important for my government and industry participants, who are worried of potential 'gotcha' journalism and bad PR, and I quickly learned when conducting interviews that a failure to adequately define clear ground rules for attribution would make the conversation much less productive.

There is a debate amongst leading ethnographers, who engage in perhaps the most sensitive and subject-centric qualitative research, as to the best practices for attribution and the anonymization of participants (for an accessible overview, see the research appendix in Nielsen, 2012). While many argue that anonymity is essential for maintaining both the trust and privacy of participants, others have argued that anonymizing subjects permits the researcher to be lazy, less careful, and less accountable when attributing statements to sources (Duneier, 2002). My specific process for discussing and obtaining consent from participants slightly evolved over the course of the research, eventually crystallizing around two questions: whether they would be willing to have their name listed in an 'appendix' in the published research (and be identified as someone that I spoke with during the course of the research), and what their stance on specific attribution of claims and arguments was. This consent was collected orally after a walkthrough of the project and its aims, or in the cases of participants who preferred written consent, via a digitally signed written consent form. If participants requested it (and a few government — and some industry — participants did), I offered to clear any direct attribution of quotes that I wished to include in this text with them.

I opted for this strategy as I believed that it struck a balance between credibility (given the centrality of data obtained via interview to my arguments, the reader is in effect required to trust that I was able to obtain access to sufficiently knowledgeable and central participants in these policy discussions; it is therefore helpful to provide a list of people who agreed to be identified as people I spoke with) and participant control. As I did not record the majority of conversations, both for technical reasons and to put participants at ease, I understood the hesitance to have quotes mis-attributed due to my shoddy notetaking, or a participant saying something candidly only to see what they said potentially inflict future reputational harm. While I

took detailed notes on each interview directly in the qualitative analysis software NVIVO, in most cases in this thesis, I did not attribute quotes from interviews directly, opting instead for (usually similar) publicly attributable quotations or general observations. In most cases, my narrative does not hinge on the identity of a specific speaker and the story they put forth.

An another important ethical discussion relates to data security and protection. I stored all project data in encrypted files and kept them secure as best as possible on a machine with full-disk encryption. I explored these ethical issues in greater depth in my ethical review application, which was approved as CUREC #SSH-DPIR-C1A-18-018.

A.1.2 Participants

What kind of people were interviewed? As mentioned above, I wished to speak to relevant stakeholders in government, industry, and civil society.

The government category is a broad one that included civil servants working in government departments and regulatory agencies, people working in the parties (political staff), and a few elected representatives. For the German case study, I interviewed a few high-level individuals, including two Members of the Bundestag working on digital policy, a former European Commissioner, the former Director General of DG Connect, and the lead on platform policy at the EU Commission. As well, I interviewed a number of advisers and staffers from all of the major German parties (with the exception of the small opposition Left party and the far-right Alternative for Germany party). The Australia and New Zealand chapter featured interviews with mid-level regulators, as well as a fantastic interview with the lead digital adviser to the Prime Minister of New Zealand, who was in charge of developing and implementing the Christchurch Call approach that is the subject of the chapter. If I had more time for all of these case studies, and perfect access, I would have liked to have interviews with key officials at the German Ministry of Justice, who refused to be interviewed, and I was unfortunately unable to get access to Gerd Billen, a key civil servant involved in the development of the NetzDG.

In Australia, I would have liked to have interviewed more political staff in the National/Liberal government, including staff within the Attorney General's Office; in New Zealand, while I spoke with folks in the Ministry of Foreign Affairs and the Department of Home Affairs, I would have liked to have interviewed more partisan political staff in the governing Labour party and perhaps the opposition.

Across the three case studies, I spoke with representatives of all of the major companies mentioned, including Facebook, Google, Twitter, and Microsoft. I also spoke with a few regional industry association representatives, who lobbied for their constituent member interests in Germany and Australia. These interviews were a mix of high-level 'policymakers' with real decision making power in the companies (for instance, I interviewed Twitter's head of global policy strategy and Facebook's global head of content regulation, both of whom work on responses to regulation around the world; I also spoke with Microsoft's Chief Digital Safety Officer, who was the chair of the Global Internet Forum to Counter Terrorism in 2019-2020) and regional policy employees on the front lines of each case study (for instance, I spoke with Facebook's head of policy for Australia, and policy managers from Google Germany, Google NZ, and Google Australia). I only conducted one interview with a Twitter employee, as both their Australian and EU/Germany representatives were unable to find the time for an interview. Throughout this process, I noticed a range of differences in industry perspectives, even within firms, and policy employees with a prior connection or sympathy for academic research were more likely to not only accept an interview request, but also be attributed on-the-record. Regional Facebook employees in particular tended to insist on keeping things totally off the record, or wished me to not even mention that the interview had taken place.

In both these cases, it was important to constantly try and unpack and unweave the statements made by interview subjects, not just as individuals but also as canny and self-interested actors often (if not always) seeking to shape the narrative around a specific policy discussion. For this reason, I drew especially on civil society interviews, as well as public documents and reporting, to triangulate the information they shared. While civil society is naturally also political, they tend to have a more

Name	Organization	Title	Date (D.M.Y)
Prabhat Agarwal	European Commission (DG Connect)	Head of Unit for Online Platforms	15.05.2020
Paul Ash	New Zealand National Cyber Policy Office	Director	15.12.2020
Ian Barber	Global Partners Digital	Legal Officer	18.12.2020
Owen Bennett	Mozilla Corporation	Internet Policy Manager	29.04.2020
Mario Brandenburg	The Free Democratic Party of Germany	Member of the Bundestag (2017—)	18.06.2020
Jordan Carter	Internet NZ	Chief Executive	9.12.2020
Tony Close	Facebook	Director of Content Regulation (Global)	20.04.2021
Barbara Docklova	Article 19	Senior Campaigner	6.05.2020
Evelyn Douek	Harvard Law School	Lecturer	30.10.2020
Alexander Fanta	Netzpolitik	Brussels Correspondent	8.04.2020
Terry Flew	University of Sydney	Professor	9.12.2020
Christiane Gillespie-Jones	Communications Alliance LTD	Director, Program Management	1.12.2020
Courtney Gregoire	Microsoft	Chief Digital Safety Officer	2.02.2021
Gabrielle Guillemin	Article 19	Digital Rights Lead	2.07.2020
	eSafety Commissioner, Australia	Senior Legal and Policy Adviser	16.02.2020
Rita Jabri-Markwell	Australian Muslim Advocacy Network	Chief Advisor	19.01.2021
Matthias Ketteman	Hans Bredow Institute	Head of Research, Online Regulation	22.04.2020
Julian King	European Commission	Commissioner for the Security Union (2016-2020)	20.05.2020
Lubos Kuklis	European Regulators Group for Audiovisual Media Services (ERGA)	Chair	16.06.2020
Lutz Mache	Google Germany	Public Policy and Government Relations Manager	30.04.2020
Josh Machin	Facebook Australia	Head of Public Policy	28.01.2021
Robert Madellin	European Commission (DG Connect)	Director General (2010-2015)	12.05.2020
Joe McNamee	European Digital Rights	Executive Director (2009-2019)	15.05.2020
Mackenzie Nelson	Algorithm Watch	Project Manager, Governing Platforms Project	28.04.2020
Paul Nemitz	European Commission (DG Justice)	Director for Fundamental Rights	5.06.2020
Matt Nguyen	Reset Australia	Policy Lead	9.02.2021
Javier Pallero	Access Now	Head of Policy	25.01.2021
Nick Pickles	Twitter	Global Head of Public Policy Strategy and Development	27.01.2021
Jörn Pohl	The Green Party of Germany	Chief of Staff, MdB Konstantin von Notz	8.05.2020
Simone Rafael	Amadeu Antonio Stiftung	Executive Director	5.06.2020
Julia Reda	European Parliament	Member of the European Parliament (2014-2018), The Greens-European FreeAlliance	27.04.2020
Alexander Ritzmann	Counter Extremism Project	Fellow	10.06.2020
Anonymous	eSafety Commissioner, Australia	Manager, Online Harms Policy	16.02.2020
Peter Thompson	Victoria University Wellington	Lecturer	24.11.2020
Mathias Schindler	European Parliament	Staffer, MEP Julia Reda's Office (2014-2018)	20.04.2020
Wolfgang Schulz	Alexander von Humboldt Institute for Internet and Society	Director	9.04.2020
Matthias Spielkamp	Algorithm Watch	Director	30.06.2020
Ellen Strickland	Internet NZ	Chief Policy Advisor, International	9.02.2021
Nicolas Suzor	Queensland University of Technology / Facebook Oversight Board	Professor / Member	3.11.2020
Mathias Vermeulen	European Parliament	Staffer, MEP Marietje Schaake (Alliance of Liberals and Democrats forEurope)	2.07.2020
Richard Wingfield	Global Partners Digital	Head of Legal	18.12.2020
Michael Woodside	Department of Internal Affairs, New Zealand	Policy Director – Gambling, Media Content, and Racing	10.03.2020
Samantha Yorke	Google Australia	Government Affairs and Public Policy Manager	19.01.2021
Ross Young	Google New Zealand	Head of Government Relations and Public Policy	26.02.2021
Jens Zimmerman	Social Democratic Party of Germany	Member of the Bundestag (2013—)	2.06.2020

Table A.2: List of Non-Anonymous Interview Participants

impartial observer perch through which to see some developments from a birds-eye-view, and were able to provide additional local, personal, and historic context.

A.2 Other Data Sources

Alongside these interviews, I drew upon the classic primary (policy documents, legal documents, memos) and secondary (journalism, reports and research prepared by academics and civil society) data sources traditionally used in social scientific desk research. For the case study chapters, I also sought to actively gain access to new primary documents that had yet been made publicly available via freedom of information access requests made to government bodies where possible. For the data chapter, I built upon a few existing data sources compiled by other researchers.

A.2.1 Freedom of Information Requests

Journalists, civil society groups, and activists have for at least a few decades actively used access to information/freedom of information requests (FOI) to obtain

various documents from, and about, governments. Nevertheless, FOI requests remain a relatively obscure qualitative method in the social sciences, despite the significant opportunities they present for researchers interested in understanding policy processes (Savage and Hyde, 2014; Walby and Luscombe, 2017). Because most internal government activity is in effect textual — involving the creation, discussion, and debate of various textual material — "Much of what is said and done in government organizations is written down or otherwise documented, and despite a range of limitations, and barriers to access, much of this material is accessible through ATI/FOI" requests (Walby and Larsen, 2012, p. 39). A small literature on the use of FOI requests as qualitative data exists, and this work highlights (a) the potential advantages to using FOI requests to bolster and triangulate interview data, especially longitudinally; (b) the new types of documents that FOI requests make available, including special process documents that include "unofficial texts that are never intended for public circulation, such as the notes and the internal memos and the emails of government employees" (Walby and Larsen, 2012, p. 33); and (c) the active role that the researcher plays as part of the FOI process of 'data production' (Walby and Luscombe, 2017), given that FOI requests require very precise wording and in some cases direct negotiation with FOI coordinators within governments.

In this thesis, I took additional advantage of the latest open-source tools for making and archiving FOI requests that have been recently developed by civil society and government transparency advocacy groups. New platforms like Avatelli provide an open source back end that have allowed for national-level FOI websites to easily pop up in multiple countries (such as FYI New Zealand, Right to Know Australia, and Ask the EU). These platforms are in effect web portals that mediate the relationship between the FOI requestee and the FOI target, seeking to make the FOI request process more accessible and externally transparent, bringing FOI requests between targets and requesters out of the realm of private communication and into an archived and searchable public format. The FOIA platforms additionally host and archive any documents provided by the FOI target, making them easily accessible to others. For this reason, these services are fantastic for research: they

are user friendly (providing a searchable directory of authorities/contact points, and a searchable directory of past requests made via the platform), are transparent (allowing others to 'see your work,' including the communication and specific language of requests which is generally hidden in a request performed via private correspondence), and offer built in archiving and citability features (each request has a URL, where others can also publicly access the relevant documents).

There is certainly a learning curve involved with making successful FOI requests. As Walby and Larsen (2012) write, the language of the request and its scope is essential as to determining its success; they advise keeping a research diary and carefully monitoring what kind of language is successful in obtaining the types of documents one wishes to obtain. Using AsktheEU and the German platform FragDenStaat, I was able to not only search for past relevant requests (and see what language had successfully yielded results; what documents were being withheld by agencies to past requests) and tweak my language accordingly. I also spoke with a few experts (a German investigative journalist who uses FOI requests daily in his work, and one of the founders of the FragDenStaat portal) to get some additional tips.

In total, I made use of seven FOIA request clusters, some of which constituted multiple requests to various government agencies. I filed two requests to the European Commission, and three to the German Ministry of Justice. I also filed a request each to various Australian government departments and the to the New Zealand Ministry of Foreign Affairs and Trade. Because New Zealand only allows citizens and permanent residents to file requests, I asked a colleague to file the request which would then be publicly archived on the FYI NZ website for me to access later. Not all these requests were successful: in one EU instance, and one Germany instance, the agency in question said that it did not have the documents I was looking for or cited a very high administrative charge for processing the request. The majority of these documents were digitized and shared simply in a reply via email, archiving them permanently on the FOI platform used to make the request; however, the two longest disclosures sent the documents separately, with the German Ministry of Justice deciding to send instead a giant shoebox filled

Actor	Department	Date	Pages	Link to Documents
EU	DG GROW, DG CNECT, DG JUST	20.04.2020	33	https://www.asktheeu.org/en/request/member_state_comments_on_netzdg#incoming-29856
EU	DG GROW	20.04.2020	0	https://www.asktheeu.org/en/request/notification_on_netzdg_amendment#incoming-26403
Germany	Ministry of Justice and Consumer Protection	09.06.2020	98	https://fragdenstaat.de/anfrage/netzdg-notifizierung/
Germany	Ministry of Justice and Consumer Protection	09.06.2020	260 (mail)	https://fragdenstaat.de/anfrage/bmjv-task-force/
Germany	Ministry of Justice and Consumer Protection	15.07.2020	0	https://fragdenstaat.de/anfrage/breife-an-internetkonzerne/
Australia	eSafety Commissioner, Attorney General's Department	08.06.2020	359	https://www.righttoknow.org.au/request/abhorrent_violent_material_act#incoming-18083
New Zealand	Ministry of Foreign Affairs and Trade	05.08.2020	33	https://fyi.org.nz/request/13466

Table A.3: FOI requests, with links to archived documents

with documents (probably intentionally, making them much more difficult for me to search and archive) and the Australian Attorney General sharing the documents via a separate email to my address (which I have uploaded as an attachment on the Right to Know page where I made my request).

A.2.2 Datasets and Coding Processes

In Chapter 3 of this thesis, I used two additional data sources: a dataset that I drew from data provided by the maintainers of the World Intermediary Liability Map project at Stanford Law School, and a new dataset I collected of public-private informal platform regulation initiatives that drew upon the coding framework established by Westerwinter (2021).

A fuller description of the Stanford data that is available has been put together by Frosio (2017), and I did not perform in-depth coding or analysis on their data, rather using it to illustrate some macro-scale longitudinal trends. However, a few words about the Westerwinter coding process and its limitations are worth mentioning here.

Firstly, the completeness of the universe: although I sought to compile my list of informal regulatory initiatives in an as systematic as possible manner, largely drawing upon existing research and civil society reporting, initiatives were inevitably missed. Following Westerwinter (2021, p. 150), I cannot claim that I have a complete universe or even a necessarily fully representative sample, given that the scale of the full population still remains unknown. Nevertheless, the data presented provide some limited, yet I think insightful, descriptive observations. Secondly, the coding strategy and the difficulty of measuring governance functions and institutional structures: as Westerwinter (2021) writes in the research appendix, their approach uses dichotomous rather than ordinal measurements:

> Measuring the design dimensions of TGIs dichotomously provides a rough proxy for capturing institutional structures. An ordinal (e.g. low, moderate, high) or continuous measure that captures for design elements not only their presence, but also the extent to which they are present (e.g. no monitoring versus self-monitoring versus peer monitoring versus independent third-party auditing), would be preferable. However, measuring institutional design intensities with validity and reliability across a large number of TGIs and across researchers is a challenging task. Dichotomous measures are less error prone and likely to produce measurements that are reproducible across coders and over time. Thus, our dichotomous measures of the institutional design features of TGIs provide sufficient detail to capture theoretically different categories of TGI design and highlight interesting empirical variation, while at the same time facilitate valid coding and minimize measurement error by focusing on clear differences across TGIs. (Westerwinter, 2021, p. 148)

In effect, I made a heuristic heuristic judgement on the governance functions and institutional structures that an initiative sought to provide/had, drawing upon publicly available documentation, as well as academic literature. The issue with the dichotomous approach is that it does not fully capture differences in degree, or in impact: for example, an initiative that only seems to fill one governance role (e.g. agenda-setting) and yet is incredibly important for agenda-setting policy debates in that broad arena will not have that qualitative difference captured in this data. Nevertheless, the data is intended simply to provide some high-level description in a more structured manner than currently exists, and to help illustrate the relative informality and lack of institutional features that characterize public-private transnational platform regulation initiatives.

References

Abbate, Janet. 2000. *Inventing the Internet*. Cambridge, MA: MIT Press. 282 pp.

Abbott, Kenneth W., Jessica F. Green, and Robert O. Keohane. 2016. "Organizational Ecology and Institutional Change in Global Governance". *International Organization* 70.2, pp. 247–277.

Abbott, Kenneth W., David Levi-Faur, and Duncan Snidal. 2017. "Theorizing Regulatory Intermediaries: The RIT Model". *The ANNALS of the American Academy of Political and Social Science* 670.1, pp. 14–35.

Abbott, Kenneth W. and Duncan Snidal. 2000. "Hard and Soft Law in International Governance". *International Organization* 54.3, pp. 421–456.

— 2009. "Strengthening International Regulation through Transnational New Governance: Overcoming the Orchestration Deficit". *Vanderbilt Journal of Transnational Law* 42, pp. 501–578.

— 2009. "The Governance Triangle: Regulatory Standards Institutions and the Shadow of the State". *The Politics of Global Regulation*. Ed. by Walter Mattli and Ngaire Woods. Princeton, NJ: Princeton University Press, pp. 44–88.

Allen, Darcy WE et al. 2021. "The Political Economy of Australian Regulatory Reform". *Australian Journal of Public Administration* 80.1, pp. 114–137.

Allen-Ebrahimian, Bethany. 2016. *The Man Who Nailed Jello to the Wall*. Foreign Policy. URL: `https://foreignpolicy.com/2016/06/29/the-man-who-nailed-jello-to-the-wall-lu-wei-china-internet-czar-learns-how-to-tame-the-web/` (visited on 05/04/2021).

Anderson, Duncan. 2017. "Film and Video Censorship in New Zealand, 1976-1994". Doctoral Dissertation. Wellington, NZ: Victoria University Wellington.

Andersson Schwarz, Jonas. 2017. "Platform Logic: An Interdisciplinary Approach to the Platform-Based Economy". *Policy & Internet* 9.4, pp. 374–394.

Andreessen, Marc. 2007. *The Three Kinds of Platforms You Meet on the Internet*. CNET. URL: `https://www.cnet.com/news/the-three-kinds-of-platforms-you-meet-on-the-internet/` (visited on 09/13/2018).

Angelopoulos, Christina and Stijn Smet. 2016. "Notice-and-Fair-Balance: How to Reach a Compromise between Fundamental Rights in European Intermediary Liability". *Journal of Media Law* 8.2, pp. 266–301.

Appleman, Bradley A. 1995. "Hate Speech: A Comparison of the Approaches Taken by the United States and Germany Notes and Comments". *Wisconsin International Law Journal* 14.2, pp. 422–439.

Ardia, David S. 2009. "Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity Under Section 230 of the Communications Decency Act". *Loyola Law Review* 43.2, pp. 373–506.

ARTICLE 19. 2018. *Germany: Responding to 'Hate Speech'*. 2018 Country Report. London, UK: ARTICLE 19.

— 2018. *Legal Analysis: Malaysian 'Anti-Fake News Act;* London, UK: ARTICLE 19.

Avant, Deborah D., Martha Finnemore, and Susan K. Sell, eds. (2010). *Who Governs the Globe?* Cambridge, UK: Cambridge University Press.

Bach, David and Abraham L. Newman. 2007. "The European Regulatory State and Global Public Policy: Micro-Institutions, Macro-Influence". *Journal of European Public Policy* 14.6, pp. 827–846.

— 2010. "Governing Lipitor and Lipstick: Capacity, Sequencing, and Power in International Pharmaceutical and Cosmetics Regulation". *Review of International Political Economy* 17.4, pp. 665–695.

Baistrocchi, Pablo Asbo. 2002. "Liability of Intermediary Service Providers in the EU Directive on Electronic Commerce". *Santa Clara Computer & High Technology Law Journal* 19.1, pp. 111–130.

Baldwin, Robert, Martin Cave, and Martin Lodge. 2012. *Understanding Regulation: Theory, Strategy, and Practice*. 2nd ed. Oxford, UK: Oxford University Press.

Barnett, Michael and Raymond Duvall. 2004. "Power in Global Governance". *Power in Global Governance*. Ed. by Michael Barnett and Raymond Duvall. Cambridge, UK: Cambridge University Press, pp. 1–32.

— 2005. "Power in International Politics". *International Organization* 59.1, pp. 39–75.

Barocas, Solon, Sophie Hood, and Malte Ziewitz. 2013. "Governing Algorithms: A Provocation Piece". Governing Algorithms: A Conference on Computation, Automation, and Control. New York University.

Barrett, Jonathan and Luke Strongman. 2012. "The Internet, the Law, and Privacy in New Zealand: Dignity with Liberty?" *International Journal of Communication* 6.1, pp. 127–143.

Bartley, Tim. 2018. *Rules without Rights: Land, Labor, and Private Authority in the Global Economy*. Oxford, UK: Oxford University Press.

Barwise, Patrick and Leo Watkins. 2018. "The Evolution of Digital Dominance: How and Why We Got to GAFA". *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple*. Ed. by Martin Moore and Damian Tambini. Oxford, UK: Oxford University Press, pp. 21–50.

Beckedahl, Markus. 2016. *Brief von Maas an Facebook*. URL: `https://fragdenstaat.de/anfrage/brief-von-maas-an-facebook/` (visited on 07/15/2020).

Benedek, Wolfgang and Matthias C. Kettemann. 2014. *Freedom of Expression and the Internet*. Strasbourg, FR: Council of Europe. 194 pp.

Benkler, Yochai. 2006. *The Wealth of Networks : How Social Production Transforms Markets and Freedom*. New Haven, CT: Yale University Press.

Bernhagen, Patrick and Neil J. Mitchell. 2010. "The Private Provision of Public Goods: Corporate Commitments and the United Nations Global Compact". *International Studies Quarterly* 54.4, pp. 1175–1187.

Berry, Jeffrey M. 2002. "Validity and Reliability Issues in Elite Interviewing". *PS: Political Science & Politics* 35.4, pp. 679–682.

Bietti, Elettra. 2021. *A Genealogy of Digital Platform Regulation*. SSRN Scholarly Paper ID 3859487. Rochester, NY: Social Science Research Network.

Bikert, Monika and Brian Fishman. 2018. *Hard Questions: What Are We Doing to Stay Ahead of Terrorists? | Facebook Newsroom*. Facebook Newsroom. URL: `https://perma.cc/YRD5-P5HU` (visited on 04/15/2019).

Bivens, Rena. 2017. "The Gender Binary Will Not Be Deprogrammed: Ten Years of Coding Gender on Facebook". *New Media & Society* 19.6, pp. 880–898.

Black, Julia. 2001. "Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a "Post-Regulatory" World". *Current Legal Problems* 54.1, pp. 103–146.
— 2008. "Constructing and Contesting Legitimacy and Accountability in Polycentric Regulatory Regimes". *Regulation & Governance* 2.2, pp. 137–164.
Blanchard, Carl. 1994. "Telecommunications Regulation in New Zealand: How Effective Is 'Light-Handed' Regulation?" *Telecommunications Policy* 18.2, pp. 154–164.
Bloch-Wehba, Hannah. 2019. "Global Platform Governance: Private Power in the Shadow of the State". *Southern Methodist University Law Review* 1, pp. 27–80.
BMFSFJ. 2016. *Hassbotschaften in Sozialen Netzwerken wirksam bekämpfen*. Bundesfamilienministerium. URL: `https://www.bmfsfj.de/bmfsfj/aktuelles/alle-meldungen/hassbotschaften-in-sozialen-netzwerken-wirksam-bekaempfen/90378` (visited on 09/14/2020).
Bogost, Ian and Nick Montfort. 2009. "Platform Studies: Frequently Questioned Answers". *Proceedings of the Digital Arts and Culture Conference*. Irvine, CA.
Börzel, Tanja A. and Thomas Risse. 2010. "Governance without a State: Can It Work?" *Regulation & Governance* 4.2, pp. 113–134.
Bowers, John and Jonathan Zittrain. 2020. "Answering Impossible Questions: Content Governance in an Age of Disinformation". *The Harvard Kennedy School (HKS) Misinformation Review* 1.1, pp. 1–8.
boyd, danah and Kate Crawford. 2012. "Critical Questions for Big Data". *Information, Communication & Society* 15.5, pp. 662–679.
boyd, danah and Nicole B. Ellison. 2007. "Social Network Sites: Definition, History, and Scholarship". *Journal of Computer-Mediated Communication* 13.1, pp. 210–230.
Bozdag, Engin and Jeroen van den Hoven. 2015. "Breaking the Filter Bubble: Democracy and Design". *Ethics and Information Technology* 17.4, pp. 249–265.
Bradford, Anu. 2012. "The Brussels Effect". *Northwestern University Law Review* 107.1, pp. 2–64.
— 2020. *The Brussels Effect: How the European Union Rules the World*. New York, NY: Oxford University Press.
Brown, Ian and Christopher Marsden. 2015. *Regulating Code*. Cambridge, MA: MIT Press.
Brühl, Tanja and Matthias Hofferberth. 2013. "Global Companies as Social Actors: Constructing Private Business in Global Governance". *The Handbook of Global Companies*. Ed. by John Mikler. Chichester, UK: John Wiley & Sons, pp. 351–370.
Bucher, Taina. 2017. "The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms". *Information, Communication & Society* 20.1, pp. 30–44.
Bucher, Taina and Anne Helmond. 2018. "The Affordances of Social Media Platforms". *The SAGE Handbook of Social Media*. Ed. by Burgess, Jean, Alice Marwick, and Thomas Poell. London, UK: SAGE, pp. 254–278.
Bulkeley, Harriet et al. 2012. "Governing Climate Change Transnationally: Assessing the Evidence from a Database of Sixty Initiatives". *Environment and Planning C: Government and Policy* 30.4, pp. 591–612.
Burgess, Jean and Nancy K. Baym. 2020. *Twitter: A Biography*. New York, NY: NYU Press.
Burgess, Jean, Alice Marwick, and Thomas Poell, eds. (2018). *The SAGE Handbook of Social Media*. London, UK: SAGE.

Burrell, Robert and Kimberlee Weatherall. 2008. "Exporting Controversy - Reactions to the Copyright Provisions of the U.S.-Australia Free Trade Agreement: Lessons for U.S. Trade Policy". *University of Illinois Journal of Law, Technology & Policy* 8.2, pp. 259–319.

Büthe, Tim. 2010. "Private Regulation in the Global Economy: A (P)Review". *Business and Politics* 12.3, pp. 1–38.

Büthe, Tim and Walter Mattli. 2011. *The New Global Rulers: The Privatization of Regulation in the World Economy*. Princeton, NJ: Princeton University Press.

Calo, Ryan and Alex Rosenblat. 2017. "The Taking Economy: Uber, Information, and Power". *Columbia Law Review* 117.6, pp. 1623–1690.

Cameron, Nadia. 2019. *Industry Warns of Negative Tech, Cultural Consequences as Social Media Violent Material Bill Passes*. CMO. URL: `https://www.cmo.com.au/article/659671/industry-warns-negative-tech-society-consequences-violent-material-bill-passed/` (visited on 10/15/2020).

Campbell, John L. 2004. *Institutional Change and Globalization*. Princeton, N.J.: Princeton University Press. 268 pp.

Caplan, Robyn. 2018. *Content or Context Moderation?* New York, NY: Data & Society Research Institute.

Carmi, Guy E. 2008. "Dignity versus Liberty: The Two Western Cultures of Free Speech". *Boston University International Law Journal* 26.2, pp. 277–374.

Carter, Jordan. 2019. *Reporting Back: InternetNZ @ the Christchurch Call in Paris*. URL: `https://internetnz.nz/blog/reporting-back-internetnz-christchurch-call-paris/` (visited on 12/08/2020).

Cashore, Benjamin William, Graeme Auld, and Deanna Newsom. 2004. *Governing through Markets: Forest Certification and the Emergence of Non-State Authority*. New Haven, CT: Yale University Press.

Castells, Manuel. 2012. *Networks of Outrage and Hope: Social Movements in the Internet Age*. Cambridge, UK: Polity. 184 pp.

Chander, Anupam. 2016. *Internet Intermediaries as Platforms for Expression and Innovation*. Waterloo, ON: Centre for International Governance Innovation.

Chandrasekharan, Eshwar et al. 2018. "The Internet's Hidden Rules: An Empirical Study of Reddit Norm Violations at Micro, Meso, and Macro Scales". *Proceedings of the ACM on Human-Computer Interaction* (CSCW), pp. 1–25.

Ciepley, David. 2013. "Beyond Public and Private: Toward a Political Theory of the Corporation". *American Political Science Review* 107.1, pp. 139–158.

Citron, Danielle Keats. 2017. "Extremist Speech, Compelled Conformity, and Censorship Creep". *Notre Dame Law Review* 93.3, pp. 1035–1072.

Citron, Danielle Keats and Benjamin Wittes. 2017. "The Internet Will Not Break: Denying Bad Samaritans Sec. 230 Immunity". *Fordham Law Review* 86.2, pp. 401–423.

Claussen, Victor. 2018. "Fighting Hate Speech and Fake News. The Network Enforcement Act (NetzDG) in Germany in the Context of European Legislation". *Rivista Di Diritto Dei Media* 2.3, pp. 1–27.

Cohen, Julie E. 2019. *Between Truth and Power: The Legal Constructions of Informational Capitalism*. Oxford, UK: Oxford University Press.

Collier, Ruth Berins, Veena Dubal, and Christopher L. Carter. 2018. "Disrupting Regulation, Regulating Disruption: The Politics of Uber in the United States". *Perspectives on Politics* 16.4, pp. 919–937.

Commonwealth of Australia. 2018. *Agency Resourcing Budget Paper No. 4, 2018-19*. Australian Treasury. URL: https://archive.budget.gov.au/2018-19/bp4/bp4.pdf (visited on 07/25/2021).

Content Works. 2019. *Social Media in Germany- The Stats You Need To Know*. Contentworks. URL: https://contentworks.agency/social-media-in-germany-the-stats-you-need-to-know/ (visited on 07/21/2021).

Corporate Europe Observatory. 2020. *Big Tech Brings out the Big Guns in Fight for Future of EU Tech Regulation*. Corporate Europe Observatory. URL: https://corporateeurope.org/en/2020/12/big-tech-brings-out-big-guns-fight-future-eu-tech-regulation (visited on 07/27/2021).

Cortell, Andrew P. and James W. Davis. 2005. "When Norms Clash: International Norms, Domestic Practices, and Japan's Internalisation of the GATT/WTO". *Review of International Studies* 31.1, pp. 3–25.

Crawford, Kate and Tarleton Gillespie. 2016. "What Is a Flag for? Social Media Reporting Tools and the Vocabulary of Complaint". *New Media & Society* 18.3, pp. 410–428.

Culpepper, Pepper D. and Kathleen Thelen. 2019. "Are We All Amazon Primed? Consumers and the Politics of Platform Power". *Comparative Political Studies* 53.2, pp. 288–318.

Dahl, Robert A. 1957. "The Concept of Power". *Behavioral Science* 2.3, pp. 201–215.

Deibert, Ronald J. et al., eds. (2008). *Access Denied: The Practice and Policy of Global Internet Filtering*. Cambridge, MA: MIT Press.

— eds. (2010). *Access Controlled: The Shaping of Power, Rights, and Rule in Cyberspace*. Cambridge, MA: MIT Press.

DeNardis, Laura. 2009. *Protocol Politics: The Globalization of Internet Governance*. Cambridge, MA: MIT Press.

Denemark, David and Shaun Bowler. 2002. "Minor Parties and Protest Votes in Australia and New Zealand: Locating Populist Politics". *Electoral Studies* 21.1, pp. 47–67.

Der Tagesspiegel. 2018. *FDP und Grüne bekräftigen Kritik am Gesetz gegen Hass im Netz*. Der Tagesspiegel. URL: https://www.tagesspiegel.de/politik/netzdg-fdp-und-gruene-bekraeftigen-kritik-am-gesetz-gegen-hass-im-netz/23816144.html (visited on 06/02/2020).

Dernbach, Andrea. 2015. *Germany Suspends Dublin Agreement for Syrian Refugees*. Euractiv. URL: https://www.euractiv.com/section/economy-jobs/news/germany-suspends-dublin-agreement-for-syrian-refugees/ (visited on 08/04/2020).

Derthick, Martha and Paul J. Quirk. 1985. *The Politics of Deregulation*. Washington, D.C.: Brookings Institution Press. 284 pp.

Deutscher Bundestag. 2017. *Bundestag beschließt Gesetz gegen strafbare Inhalte im Internet*. Deutscher Bundestag. URL: https://perma.cc/26CA-V3TW (visited on 09/14/2020).

Diamond, Larry. 2010. "Liberation Technology". *Journal of Democracy* 21.3, pp. 69–83.

Dingwerth, Klaus and Philipp Pattberg. 2006. "Global Governance as a Perspective on World Politics". *Global Governance* 12.2, pp. 185–203.

Dingwerth, Klaus and Philipp Pattberg. 2009. "World Politics and Organizational Fields: The Case of Transnational Sustainability Governance". *European Journal of International Relations* 15.4, pp. 707–743.

Dostal, Jörg Michael. 2015. "The Pegida Movement and German Political Culture: Is Right-Wing Populism Here to Stay?" *The Political Quarterly* 86.4, pp. 523–531.

Douek, Evelyn. 2020. "Australia's 'Abhorrent Violent Material' Law: Shouting 'Nerd Harder' and Drowning Out Speech". *Australian Law Journal* 94, pp. 41–60.

— 2020. *The Rise of Content Cartels*. Columbia University: Knight First Amendment Institute.

Drezner, Daniel W. 2008. *All Politics Is Global: Explaining International Regulatory Regimes*. Princeton, NJ: Princeton University Press. 265 pp.

Duckett, Chris. 2019. *Australia's Abhorrent Video Streaming Legislation Rammed through Parliament*. ZDNet. URL: `https://www.zdnet.com/article/australian-abhorrent-video-streaming-legislation-rammed-through-senate/` (visited on 12/08/2020).

Duguay, Stefanie, Jean Burgess, and Nicolas Suzor. 2018. "Queer Women's Experiences of Patchwork Platform Governance on Tinder, Instagram, and Vine". *Convergence* 26.2, pp. 237–252.

Duneier, Mitchell. 2002. "What Kind of Combat Sport Is Sociology?" *American Journal of Sociology* 107.6, pp. 1551–1576.

Dunkley, Daniel. 2021. *Twitter Officially Enters New Zealand; Seeks Developers*. Business Desk NZ. URL: `https://businessdesk.co.nz/article/media/twitter-officially-enters-new-zealand-seeks-developers` (visited on 07/24/2021).

Easton, Brian. 1997. *The Commercialisation of New Zealand*. Auckland, NZ: Auckland University Press. 300 pp.

Eberlein, Burkard and Claudio M. Radaelli. 2010. "Mechanisms of Conflict Management in EU Regulatory Policy". *Public Administration* 88.3, pp. 782–799.

Echikson, William and Olivia Knodt. 2018. *Germany's NetzDG: A Key Test for Combatting Online Hate*. SSRN Scholarly Paper ID 3300636. Rochester, NY: Social Science Research Network.

Edelman, Benjamin G. and Damien Geradin. 2015. "Efficiencies and Regulatory Shortcuts: How Should We Regulate Companies like Airbnb and Uber". *Stanford Technology Law Review* 19.2, pp. 293–328.

Elers, Christine Helen and Pooja Jayan. 2020. ""This Is Us": Free Speech Embedded in Whiteness, Racism and Coloniality in Aotearoa, New Zealand". *First Amendment Studies* 54.2, pp. 236–249.

Erickson, Kristofer and Martin Kretschmer. 2018. "This Video Is Unavailable: Analyzing Copyright Takedown of User-Generated Content on YouTube". *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 9.1, pp. 75–89.

Euchner, Eva-Maria. 2020. "Ruling under a Shadow of Moral Hierarchy: Regulatory Intermediaries in the Governance of Prostitution". *Regulation & Governance* Early View, pp. 1–22.

European Commission. 2002. *12 EU Countries Miss E-Commerce Directive Implementation Deadline*. CORDIS. URL: `https://cordis.europa.eu/article/id/17873-12-eu-countries-miss-ecommerce-directive-implementation-deadline` (visited on 07/21/2021).

— 2016. *European Commission and IT Companies Announce Code of Conduct on Illegal Online Hate Speech*. URL: `https://perma.cc/3M7U-5AQY`.

— 2020. *Proposal for a Regulation of the European Parliament and of the Council on Contestable and Fair Markets in the Digital Sector (Digital Markets Act)*. URL: https://ec.europa.eu/info/sites/default/files/proposal-regulation-single-market-digital-services-digital-services-act_en.pdf (visited on 07/21/2021).

Evans, David S., Andrei Hagiu, and Richard Schmalensee. 2008. *Invisible Engines: How Software Platforms Drive Innovation and Transform Industries*. Cambridge MA: MIT Press.

Evans, Peter C. and Annabelle Gawer. 2016. *The Rise of the Platform Enterprise: A Global Survey*. Report. University of Surrey, UK: The Center for Global Enterprise.

Eynon, Rebecca, Ralph Schroeder, and Jenny Fry. 2009. "New Techniques in Online Research: Challenges for Research Ethics". *Twenty-First Century Society* 4.2, pp. 187–199.

Fabo, Brian et al. 2017. *An Overview of European Platforms: Scope and Business Models*. Brussels, BE: European Commission Joint Research Centre.

Falkner, Robert. 2017. *Business Power and Conflict in International Environmental Politics*. Cham, CH: Springer.

Farrell, Henry. 2006. "Regulating Information Flows: States, Private Actors, and E-Commerce". *Annual Review of Political Science* 9.1, pp. 353–374.

Farrell, Henry and Abraham L. Newman. 2010. "Making Global Markets: Historical Institutionalism in International Political Economy". *Review of International Political Economy* 17.4, pp. 609–638.

— 2014. "Domestic Institutions beyond the Nation-State: Charting the New Interdependence Approach". *World Politics* 66.2, pp. 331–363.

— 2015. "The New Politics of Interdependence: Cross-National Layering in Trans-Atlantic Regulatory Disputes". *Comparative Political Studies* 48.4, pp. 497–526.

— 2016. "The New Interdependence Approach: Theoretical Development and Empirical Demonstration". *Review of International Political Economy* 23.5, pp. 713–736.

— 2018. "Linkage Politics and Complex Governance in Transatlantic Surveillance". *World Politics* 70.4, pp. 515–554.

— 2019. *Of Privacy and Power: The Transatlantic Struggle Over Freedom and Security*. Princeton, NJ: Princeton University Press.

— 2019. "Weaponized Interdependence: How Global Economic Networks Shape State Coercion". *International Security* 44.1, pp. 42–79.

Fiedler, Kirsten. 2015. *EU Internet Forum – behind Closed Doors and without Civil Society*. EDRi. URL: https://edri.org/launch-of-the-eu-internet-forum-behind-closed-doors-and-without-civil-society/ (visited on 06/04/2020).

Fiesler, Casey et al. 2018. "Reddit Rules! Characterizing an Ecosystem of Governance". *Twelfth International AAAI Conference on Web and Social Media*.

Finck, Michèle. 2017. "Digital Co-Regulation: Designing a Supranational Legal Framework for the Platform Economy". *LSE Law, Society and Economy Working Paper* 15, pp. 1–30.

Fioretos, Orfeo. 2011. "Historical Institutionalism in International Relations". *International Organization* 65.2, pp. 367–399.

Fisher, Adam. 2018. *Sex, Beer, and Coding: Inside Facebook's Wild Early Days in Palo Alto*. Wired Magazine. URL: https://www.wired.com/story/sex-beer-and-coding-inside-facebooks-wild-early-days/ (visited on 08/02/2018).

Flew, Terry and Rosalie Gillett. 2020. *Platform Policy: Evaluating Different Responses to the Challenges of Platform Power*. SSRN Scholarly Paper ID 3628959. Rochester, NY: Social Science Research Network.

Flew, Terry and Derek Wilding. 2020. "The Turn to Regulation in Digital Communication: The ACCC's Digital Platforms Inquiry and Australian Media Policy". *Media, Culture & Society* 43.1, pp. 48–65.

Flew, Terry et al. 2021. "Return of the Regulatory State: A Stakeholder Analysis of Australia's Digital Platforms Inquiry and Online News Policy". *The Information Society* 37.2, pp. 128–145.

Fraag den Staat. 2017. *Task Force „Umgang Mit Rechtswidrigen Hassbotschaften Im Internet"*. URL: `https://fragdenstaat.de/anfrage/task-force-umgang-mit-rechtswidrigen-hassbotschaften-im-internet/` (visited on 04/07/2020).

Fransen, Luc W. 2012. "Multi-Stakeholder Governance and Voluntary Programme Interactions: Legitimation Politics in the Institutional Design of Corporate Social Responsibility". *Socio-Economic Review* 10.1, pp. 163–192.

Fransen, Luc W. and Ans Kolk. 2007. "Global Rule-Setting for Business: A Critical Analysis of Multi-Stakeholder Standards". *Organization* 14.5, pp. 667–684.

Freelon, Deen, Charlton McIlwain, and Meredith Clark. 2018. "Quantifying the Power and Consequences of Social Media Protest". *New Media & Society* 20.3, pp. 990–1011.

Frosio, Giancarlo F. 2017. "Internet Intermediary Liability: WILMap, Theory and Trends". *Indian Journal of Law and Technology* 13, pp. 1–16.

— 2018. "Why Keep a Dog and Bark Yourself? From Intermediary Liability to Responsibility". *International Journal of Law and Information Technology* 26.1, pp. 1–33.

Fuchs, Doris. 2013. "Theorizing the Power of Global Companies". *The Handbook of Global Companies*. Ed. by John Mikler. New York: Wiley, pp. 77–95.

Fuchs, Doris A. 2007. *Business Power in Global Governance*. Boulder, CO: Lynne Rienner.

Fujimura, Naofumi. 2016. "Re-Election Isn't Everything: Legislators' Goal-Seeking and Committee Activity in Japan". *The Journal of Legislative Studies* 22.2, pp. 153–174.

Fukuyama, Francis. 2013. "What Is Governance?" *Governance* 26.3, pp. 347–368.

Fukuyama, Francis and Andrew Grotto. 2020. "Comparative Media Regulation in the United States and Europe". *Social Media and Democracy: The State of the Field and Prospects for Reform*. Ed. by Nathaniel Persily and Joshua Tucker. Cambridge, UK: Cambridge University Press, pp. 199–219.

Garden, Charlotte. 2016. "The Deregulatory First Amendment at Work". *Harvard Civil Rights-Civil Liberties Law Review* 51, p. 323.

Garton Ash, Timothy, Robert Gorwa, and Danaë Metaxa. 2019. *Glasnost! Nine Ways Facebook Can Make Itself a Better Forum for Free Speech and Democracy*. Oxford, UK: Reuters Institute for the Study of Journalism.

Gathmann, Florian. 2015. *Heidenau: Sigmar Gabriel besucht Flüchtlingsunterkunft*. Der Spiegel. URL: `https://www.spiegel.de/politik/deutschland/heidenau-sigmar-gabriel-besucht-fluechtlingsunterkunft-a-1049582.html` (visited on 06/11/2020).

Gathmann, Florian and Horand Knaup. 2017. *Heiko Maas zu Machtanspruch von Martin Schulz: "Alles andere wäre armselig"*. Der Spiegel. URL: `https://www.spiegel.de/politik/deutschland/heiko-maas-zu-machtanspruch-von-`

martin-schulz-alles-andere-waere-armselig-a-1139270.html (visited on 08/03/2020).

Gauja, Anika, Marian Sawer, and Marian Simms, eds. (2020). *Morrison's Miracle: The 2019 Australian Federal Election*. Canberra, AU: ANU Press.

Gawer, Annabelle, ed. (2011). *Platforms, Markets and Innovation*. Cheltenham, UK: Edward Elgar Publishing. 413 pp.

Gawer, Annabelle and Michael A. Cusumano. 2014. "Industry Platforms and Ecosystem Innovation". *Journal of Product Innovation Management* 31.3, pp. 417–433.

Gelber, Katharine and Luke McNamara. 2013. "Freedom of Speech and Racial Vilification in Australia: 'The Bolt Case' in Public Discourse". *Australian Journal of Political Science* 48.4, pp. 470–484.

Gerring, John. 2004. "What Is a Case Study and What Is It Good For?" *American Political Science Review* 98.2, pp. 341–354.

— 2006. *Case Study Research: Principles and Practices*. Cambridge, UK: Cambridge University Press.

Gillespie, Tarleton. 2010. "The Politics of 'Platforms'". *New Media & Society* 12.3, pp. 347–364.

— 2014. *Media Technologies: The Relevance of Algorithms*. Cambridge, MA: MIT Press.

— 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT: Yale University Press.

— 2018. "Regulation of and by Platforms". *The SAGE Handbook of Social Media*. Ed. by Burgess, Jean, Alice Marwick, and Thomas Poell. London, UK: SAGE, pp. 254–278.

Gillespie, Tarleton and Nick Seaver. 2015. *Critical Algorithm Studies: A Reading List*. Microsoft Research, New England: Social Media Collective.

Gillespie, Tarleton et al. 2020. "Expanding the Debate about Content Moderation: Scholarly Research Agendas for the Coming Policy Debates". *Internet Policy Review* 9.4, pp. 1–29.

Given, Jock. 2004. "'Not Unreasonably Denied': Australian Content after AUSFTA". *Media International Australia* 111.1, pp. 8–22.

Glaser, Stefan et al. 2008. *Protection of Minors on the Internet: Jungendschutz Annual Report*. Mainz, DE: Jungendschutz.

Goldsmith, Ben and Julian Thomas. 2012. "The Convergence Review and the Future of Australian Content Regulation". *Australian Journal of Telecommunications and the Digital Economy* 62.3, pp. 44–1.

Goldsmith, Jack L and Tim Wu. 2006. *Who Controls the Internet?: Illusions of a Borderless World*. New York, NY: Oxford University Press.

Gollatz, Kirsten and Leontine Jenner. 2018. *Hate Speech and Fake News - How Two Concepts Got Intertwined and Politicised | Digital Society Blog*. HIIG. URL: https://www.hiig.de/en/hate-speech-fake-news-two-concepts-got-intertwined-politicised/ (visited on 09/10/2020).

Gorwa, Robert. 2019. "The Platform Governance Triangle: Conceptualising the Informal Regulation of Online Content". *Internet Policy Review* 8.2, pp. 1–18.

— 2019. "What Is Platform Governance?" *Information, Communication & Society* 22.6, pp. 854–871.

— 2021. "Elections, Institutions, and the Regulatory Politics of Platform Governance: The Case of the German NetzDG". *Telecommunications Policy* 45.6, p. 102145.

Gorwa, Robert, Reuben Binns, and Christian Katzenbach. 2020. "Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance". *Big Data & Society* 7.1, p. 205395171989794.

Gorwa, Robert and Timothy Garton Ash. 2020. "Democratic Transparency in the Platform Society". *Social Media and Democracy: The State of the Field and Prospects for Reform*. Ed. by Nathaniel Persily and Joshua Tucker. Cambridge, UK: Cambridge University Press, pp. 286–312.

Gorwa, Robert and Anton Peez. 2020. "Big Tech Hits the Diplomatic Circuit: Norm Entrepreneurship, Policy Advocacy, and Microsoft's Cybersecurity Tech Accord". *Governing Cyberspace: Behaviour, Power, and Diplomacy*. Ed. by Dennis Broeders and Bibi van den Berg. Lanham, MD: Rowman & Littlefield, pp. 263–284.

Graham-McLay, Charlotte. 2020. *'I'm the Last Censor in the Western World': New Zealand's David Shanks Tackles the c-Word*. The Guardian. URL: `http://www.theguardian.com/world/2020/jan/10/im-the-last-censor-in-the-western-world-new-zealands-david-shanks-tackles-the-c-word` (visited on 06/26/2021).

Green, Jessica F. 2014. *Rethinking Private Authority: Agents and Entrepreneurs in Global Environmental Governance*. Princeton, NJ: Princeton University Press.

Green, Jessica F. and Graeme Auld. 2017. "Unbundling the Regime Complex: The Effects of Private Authority". *Transnational Environmental Law* 6.2, pp. 259–284.

Greenleaf, Graham. 2019. *Global Tables of Data Privacy Laws and Bills*. SSRN Scholarly Paper ID 3380794. Rochester, NY: Social Science Research Network.

Grimmelmann, James. 2008. "The Google Dilemma". *New York Law School Law Review* 53.1, pp. 939–960.

— 2015. "The Virtues of Moderation". *Yale Journal of Law & Technology* 17.1, pp. 43–109.

Haggart, Blayne. 2014. *Copyfight: The Global Politics of Digital Copyright Reform*. Toronto, CA: University of Toronto Press.

Haggart, Blayne and Clara Iglesias Keller. 2021. "Democratic Legitimacy in Global Platform Governance". *Telecommunications Policy* 45.6, p. 102152.

Hall, Rodney Bruce and Thomas J. Biersteker, eds. (2002a). *The Emergence of Private Authority in Global Governance*. Cambridge, UK: Cambridge University Press. 276 pp.

— 2002. "The Emergence of Private Authority in the International System". *The Emergence of Private Authority in Global Governance*. Ed. by Rodney Bruce Hall and Thomas J. Biersteker. Cambridge, UK: Cambridge University Press, pp. 3–22.

Halliday, Josh. 2012. *Twitter's Tony Wang: 'We Are the Free Speech Wing of the Free Speech Party'*. The Guardian. URL: `http://www.theguardian.com/media/2012/mar/22/twitter-tony-wang-free-speech` (visited on 07/15/2021).

Hallin, Daniel C. and Paolo Mancini. 2004. *Comparing Media Systems: Three Models of Media and Politics*. Cambridge, UK: Cambridge University Press.

Hammond, Thomas H. and Jack H. Knott. 1988. "The Deregulatory Snowball: Explaining Deregulation in the Financial Industry". *The Journal of Politics* 50.1, pp. 3–30.

Hansen, Suella and Noelle Jones. 2017. "New Zealand Telecommunications: The Actual Situation-Legislation and Regulations". *Australian Journal of Telecommunications and the Digital Economy* 5.3, pp. 83–88.

Hargittai, Eszter. 2007. "The Social, Political, Economic, and Cultural Dimensions of Search Engines: An Introduction". *Journal of Computer-Mediated Communication* 12.3, pp. 769–777.

Harrison, Joel. 2006. "Truth, Civility, and Religious Battlegrounds: The Contest between Religious Vilification Laws and Freedom of Expression". *Te Mata Koi: Auckland University Law Review* 12, pp. 71–96.

Hartlapp, Miriam, Julia Metz, and Christian Rauh. 2014. *Which Policy for Europe?: Power and Conflict inside the European Commission*. Oxford, UK: Oxford University Press.

Hassib, Bassant and James Shires. 2021. "Manipulating Uncertainty: Cybersecurity Politics in Egypt". *Journal of Cybersecurity* 7.1, pp. 1–16.

He, Danya. 2020. "Governing Hate Content Online: How the Rechtsstaat Shaped the Policy Discourse on the NetzDG in Germany". *International Journal of Communication* 14, pp. 3746–3768.

Helberger, Natali. 2020. "The Political Power of Platforms: How Current Attempts to Regulate Misinformation Amplify Opinion Power". *Digital Journalism* 8.6, pp. 842–854.

Heldt, Amélie. 2019. "Reading between the Lines and the Numbers: An Analysis of the First NetzDG Reports". *Internet Policy Review* 8.2, pp. 1–18.

— 2020. *Germany Is Amending Its Online Speech Act NetzDG... but Not Only That*. Internet Policy Review. URL: `https://policyreview.info/articles/news/germany-amending-its-online-speech-act-netzdg-not-only/1464` (visited on 04/06/2020).

Hellner, Michael. 2004. "Country of Origin Principle in the E-Commerce Directive-A Conflict with Conflict of Laws?" *European Review of Private Law* 12.2, pp. 193–213.

Helmond, Anne. 2015. "The Platformization of the Web: Making Web Data Platform Ready". *Social Media + Society* 1.2, p. 2056305115603080.

Herman, Bill D. 2013. *The Fight over Digital Rights: The Politics of Copyright and Technology*. Cambridge, UK: Cambridge University Press.

Hernández, Tanya Katerí. 2010. "Hate Speech and the Language of Racism in Latin America: A Lens for Reconsidering Global Hate Speech Restrictions and Legislation Models". *University of Pennsylvania Journal of International Law* 32, pp. 805–841.

Hinger, Sophie. 2016. "Asylum in Germany: The Making of the 'Crisis' and the Role of Civil Society". *Human Geography* 9.2, pp. 78–88.

Hofferberth, Matthias, ed. (2019). *Corporate Actors in Global Governance: Business as Usual or New Deal?* Boulder, CO: Lynne Rienner. 288 pp.

Hoffmann, Anna Lauren, Nicholas Proferes, and Michael Zimmer. 2018. "'Making the World More Open and Connected': Mark Zuckerberg and the Discursive Construction of Facebook and Its Users". *New Media & Society* 20.1, pp. 199–218.

Hoffmann-Riem, Wolfgang. 2016. "Selbstregelung, Selbstregulierung Und Regulierte Selbstregulierung Im Digitalen Kontext". *Neue Macht- Und Verantwortungsstrukturen in Der Digitalen Welt*. Ed. by Michael Fehling and Utz Schliesky. Baden-Baden, DE: Nomos, pp. 27–52.

Holmes, Kyle, Mark Balnaves, and Yini Wang. 2015. "Red Bags and WeChat (Wēixìn): Online Collectivism during Massive Chinese Cultural Events". *Global Media Journal* 9.1, pp. 15–26.

Hunter, Fergus and Jennifer Duke. 2019. *Facebook Censured by Government for Failure to Act on Livestreaming Concerns*. The Sydney Morning Herald. URL: `https:`

//www.smh.com.au/politics/federal/facebook-censured-by-government-for-failure-to-act-on-livestreaming-concerns-20190326-p517sb.html (visited on 03/19/2021).

Huszti-Orban, Krisztina. 2017. "Countering Terrorism and Violent Extremism Online: What Role for Social Media Platforms?" *Platform Regulations: How Platforms Are Regulated and How They Regulate Us*. Ed. by Luca Belli and Nicolo Zingales. Rio de Janeiro: Fundação Getulio Vargas, pp. 189–212.

Jacob, Lisa. 2018. *87% of Germans Approve of Social Media Regulation Law*. Dalia Research. URL: https://daliaresearch.com/blog/blog-germans-approve-of-social-media-regulation-law/ (visited on 07/21/2021).

Jennen, Birgit and Ania Nussbaum. 2021. *Germany and France Oppose Trump's Twitter Exile*. Bloomberg. URL: https://www.bloomberg.com/news/articles/2021-01-11/merkel-sees-closing-trump-s-social-media-accounts-problematic (visited on 07/15/2021).

Jørgensen, Rikke Frank, ed. (2019). *Human Rights in the Age of Platforms*. Cambridge, MA: MIT Press.

Jouanjan, Olivier. 2009. "Freedom of Expression in the Federal Republic of Germany Symposium: An Ocean Apart - Freedom of Expression in Europe and the United States". *Indiana Law Journal* 84.3, pp. 867–884.

Julià-Barceló, Rosa and Kamiel J Koelman. 2000. "Intermediary Liability in the E-Commerce Directive: So Far So Good, But It's Not Enough". *Computer Law & Security Review* 16.4, pp. 231–239.

Jungendschutz. 2016. *Ergebnisse Des Monitorings von Beschwerdemechanismen Jugendaffiner Dienste*. URL: https://perma.cc/5JQU-FQPV (visited on 09/14/2020).

Jupille, Joseph, Walter Mattli, and Duncan Snidal. 2017. "Dynamics of Institutional Choice". *International Politics and Institutions in Time*. Ed. by Orfeo Fioretos. Oxford, UK: Oxford University Press, pp. 118–143.

Kalyanpur, Nikhil and Abraham L. Newman. 2019. "The MNC-Coalition Paradox: Issue Salience, Foreign Firms and the General Data Protection Regulation". *JCMS: Journal of Common Market Studies* 57.3, pp. 448–467.

Karpen, Ulrich, Nils Molle, and Simon Schwarz. 2007. "Freedom of Expression and the Administration of Justice in Germany". *European Journal of Law Reform* 9.1, pp. 63–90.

Karpf, David. 2017. "Digital Politics After Trump". *Annals of the International Communication Association* 41.2, pp. 198–207.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2010. *The Worldwide Governance Indicators: Methodology and Analytical Issues*. World Bank Policy Research Working Paper 5430. New York, NY.

Kaye, David. 2019. *Speech Police: The Global Struggle to Govern the Internet*. New York, NY: Columbia Global Reports.

Kaye, David and Fionnuala Ní Aoláin. 2019. *Letter from UN Rapporteurs to the Australian Minister for Foreign Affairs*. Freedex. URL: https://freedex.org/wp-content/blogs.dir/2015/files/2019/04/OL-AUS-04.04.19-5.2019-2.pdf (visited on 06/08/2020).

Keck, Margaret E. and Kathryn Sikkink. 2014. *Activists beyond Borders: Advocacy Networks in International Politics*. Ithaca, NY: Cornell University Press.

Keller, Daphne. 2018. "The Right Tools: Europe's Intermediary Liability Laws and the EU 2016 General Data Protection Regulation". *Berkeley Technology Law Journal* 33, pp. 287–364.
Keller, Eileen. 2018. "Noisy Business Politics: Lobbying Strategies and Business Influence after the Financial Crisis". *Journal of European Public Policy* 25.3, pp. 287–306.
Kellerman, Miles. 2019. "The Proliferation of Multilateral Development Banks". *The Review of International Organizations* 14.1, pp. 107–145.
Kello, Lucas. 2017. *The Virtual Weapon and International Order*. New Haven, CT: Yale University Press.
Kelly, Makena. 2020. *Joe Biden Wants to Revoke Section 230*. The Verge. URL: `https://www.theverge.com/2020/1/17/21070403/joe-biden-president-election-section-230-communications-decency-act-revoke` (visited on 08/22/2021).
Kelsey, Jane. 1993. *Rolling Back the State: Privatisation of Power in Aotearoa/New Zealand*. Wellington, NZ: B. Williams Books.
Kenny, Katie. 2019. *Chief Censor David Shanks Says an Entirely New Media Regulator May Be Needed*. Stuff. URL: `https://www.stuff.co.nz/technology/digital-living/116776465/chief-censor-david-shanks-says-an-entirely-new-media-regulator-may-be-needed` (visited on 06/26/2021).
Kettemann, Matthias C. 2020. *The Normative Order of the Internet*. New York, NY: Oxford University Press.
Kettemann, Matthias C. and Wolfgang Schulz. 2020. *Setting Rules for 2.7 Billion: A (First) Look into Facebook's Norm-Making System*. Hamburg, DE: Hans-Bredow-Institut.
Khan, Lina M. 2017. "Amazon's Antitrust Paradox". *Yale Law Journal* 126.3, pp. 712–805.
Khan, Lina M. and Sandeep Vaheesan. 2017. "Market Power and Inequality: The Antitrust Counterrevolution and Its Discontents". *Harvard Law & Policy Review* 11.2, pp. 235–294.
Kirschbaum, Erik. 2015. *German Justice Minister Takes Aim at Facebook over Racist Posts*. Reuters. URL: `https://www.reuters.com/article/us-facebook-germany-racism-idUSKCN0QW1SG20150827` (visited on 06/11/2020).
Kleine, Mareike. 2013. *Informal Governance in the European Union: How Governments Make International Organizations Work*. Ithaca, NY: Cornell University Press.
Klonick, Kate. 2017. "The New Governors: The People, Rules, and Processes Governing Online Speech". *Harvard Law Review* 131.6, pp. 1598–1670.
— 2020. "The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression". *Yale Law Journal* 129.8, pp. 2418–2499.
Kohl, Uta. 2007. *Jurisdiction and the Internet: Regulatory Competence over Online Activity*. Cambridge, UK: Cambridge University Press.
— 2012. "The Rise and Rise of Online Intermediaries in the Governance of the Internet and beyond – Connectivity Intermediaries". *International Review of Law, Computers & Technology* 26.2-3, pp. 185–210.
Kosseff, Jeff. 2016. "The Gradual Erosion of the Law That Shaped the Internet: Section 230's Evolution over Two Decades". *Columbia Science and Technology Law Review* 18.1, pp. 1–41.
— 2019. "First Amendment Protection for Online Platforms". *Computer Law & Security Review* 35.5, p. 105340.

Kosseff, Jeff. 2019. *The Twenty-Six Words That Created the Internet*. Ithaca, NY: Cornell University Press. 326 pp.

Krahulcova, Lucie and Brett Solomon. 2019. *Australia's Plans for Internet Regulation: Aimed at Terrorism, but Harming Human Rights*. Access Now. URL: https://www.accessnow.org/australias-plans-to-regulate-social-media-bound-to-boomerang/ (visited on 11/26/2020).

Krauss, Martin. 2015. *Null Toleranz bei Hassparolen*. Jüdische Allgemeine. URL: https://www.juedische-allgemeine.de/politik/null-toleranz-bei-hassparolen/ (visited on 07/15/2020).

Kraut, Robert E. and Paul Resnick. 2012. *Building Successful Online Communities: Evidence-Based Social Design*. Cambridge, MA: MIT Press.

Kreiss, Daniel. 2012. *Taking Our Country Back: The Crafting of Networked Politics from Howard Dean to Barack Obama*. Oxford, UK: Oxford University Press.

Krishnamurthy, Vivek and Jessica Fjeld. 2020. *CDA 230 Goes North American? Examining the Impacts of the USMCA's Intermediary Liability Provisions in Canada and the United States*. SSRN Scholarly Paper ID 3645462. Rochester, NY: Social Science Research Network.

Krotoszynski, Ronald J. 2006. *The First Amendment in Cross-Cultural Perspective: A Comparative Legal Analysis of the Freedom of Speech*. New York, NY: NYU Press. 318 pp.

Kuczerawy, Aleksandra. 2015. "Intermediary Liability & Freedom of Expression: Recent Developments in the EU Notice & Action Initiative". *Computer Law & Security Review* 31.1, pp. 46–56.

— 2018. *Intermediary Liability and Freedom of Expression in the EU: From Concepts to Safeguards*. Cambridge, UK: Intersentia.

Kuczerawy, Aleksandra and Jef Ausloos. 2015. "From Notice-and-Takedown to Notice-and-Delist: Implementing Google Spain". *Colorado Tech Law Journal* 14.2, pp. 219–258.

Kumarasingham, Harshan. 2013. "Exporting Executive Accountability? Westminster Legacies of Executive Power". *Parliamentary Affairs* 66.3, pp. 579–596.

Langlois, Ganaele. 2013. "Participatory Culture and the New Governance of Communication: The Paradox of Participatory Media". *Television & New Media* 14.2, pp. 91–105.

Leech, Beth L. 2002. "Interview Methods in Political Science". *Political Science & Politics* 35.04, pp. 663–664.

Lessig, Lawrence. 2009. *Code: And Other Laws of Cyberspace*. 2nd ed. New York: Basic Books.

Liesching, Marc. 2017. *NetzDG-Entwurf basiert auf Bewertungen von Rechtslaien*. beck-community. URL: https://community.beck.de/2017/05/26/netzdg-entwurf-basiert-auf-bewertungen-von-rechtslaien (visited on 05/14/2020).

Llansó, Emma. 2016. *Takedown Collaboration by Private Companies Creates Troubling Precedent*. Center for Democracy & Technology. URL: https://cdt.org/blog/takedown-collaboration-by-private-companies-creates-troubling-precedent/ (visited on 04/16/2019).

— 2020. *Human Rights NGOs in Coalition Letter to GIFCT*. Center for Democracy and Technology. URL: https://cdt.org/insights/human-rights-ngos-in-coalition-letter-to-gifct/ (visited on 08/05/2020).

Logan, Lucas. 2019. "Current Policy Issues in Internet Intermediary Liability". *Oxford Research Encyclopedia of Communication*, pp. 1–21.

LVZ. 2019. *Leipziger Staatsrechtler klagt im Namen der FDP gegen das NetzDG*. Leipziger Volkszeitung. URL: https://www.lvz.de/Region/Mitteldeutschland/Leipziger-Staatsrechtler-klagt-im-Namen-der-FDP-gegen-das-NetzDG (visited on 06/18/2020).

Lynch, Jenna. 2019. *Jacinda Ardern Will Not Follow Australia's Hard-Line Response to Extremist Content*. Newshub. URL: https://www.newshub.co.nz/home/politics/2019/07/jacinda-ardern-will-not-follow-australia-s-hard-line-response-to-extremist-content.html (visited on 06/08/2020).

Lyons, Kim. 2020. *Mark Zuckerberg 'Worried' about China's Influence on Internet Regulation*. The Verge. URL: https://www.theverge.com/2020/5/18/21262707/zuckerberg-china-regulation-privacy-facebook (visited on 07/15/2021).

MacKinnon, Rebecca. 2013. *Consent of the Networked: The Worldwide Struggle for Internet Freedom*. New York, NY: Basic Books.

MacKinnon, Rebecca and Roya Pakzad. 2018. *Private Sector Roles and Responsibilities: Protecting Quality of Discourse, Diversity of Content and Civic Engagement on Digital Platforms and Social Media*. Waterloo, ON: Centre for International Governance Innovation.

Maclay, Colin Miles. 2010. "Protecting Privacy and Expression Online: Can the Global Network Initiative Embrace the Character of the Net". *Access Controlled: The Shaping of Power, Rights, and Rule in Cyberspace*. Ed. by Ronald J. Deibert et al. Cambridge, MA: MIT Press, pp. 87–108.

Majone, Giandomenico. 1998. "Public Policy and Administration: Ideas, Interestsand Institutions". *A New Handbook Of Political Science*. Ed. by Robert E. Goodin and Hans-Dieter Klingemann. Oxford, UK: Oxford University Press, p. 687.

— 1999. "The Regulatory State and Its Legitimacy Problems". *West European Politics* 22.1, pp. 1–24.

Mann, Monique, Angela Daly, and Adam Molnar. 2020. "Regulatory Arbitrage and Transnational Surveillance: Australia's Extraterritorial Assistance to Access Encrypted Communications". *Internet Policy Review* 9.3, pp. 1–20.

March, James G. and Johan P. Olsen. 2011. "The Logic of Appropriateness". *The Oxford Handbook of Political Science*. Ed. by Robert E. Goodin. Oxford, UK: Oxford University Press, pp. 479–497.

Maréchal, Nathalie. 2015. "Ranking Digital Rights: Human Rights, the Internet and the Fifth Estate". *International Journal of Communication* 9.1, pp. 3440–3449.

Maréchal, Nathalie and Sarah T. Roberts. 2018. *Researching ICT Companies: A Field Guide for Civil Society Researchers*. University of Pennslyvania: Internet Policy Observatory.

Markham, Annette, Elizabeth Buchanan, and AoIR Ethics Working Committee. 2012. *Ethical Decision-Making and Internet Research: Version 2.0*. Association of Internet Researchers Ethics Committee Report.

Marsden, Christopher. 2011. *Internet Co-Regulation: European Law, Regulatory Governance and Legitimacy in Cyberspace*. Cambridge, UK: Cambridge University Press.

Mason-Bish, Hannah. 2019. "The Elite Delusion: Reflexivity, Identity and Positionality in Qualitative Research". *Qualitative Research* 19.3, pp. 263–276.

Matamoros-Fernández, Ariadna. 2017. "Platformed Racism: The Mediation and Circulation of an Australian Race-Based Controversy on Twitter, Facebook and YouTube". *Information, Communication & Society* 20.6, pp. 930–946.

Matias, J. Nathan. 2019. "The Civic Labor of Volunteer Moderators Online". *Social Media + Society* 5.2, p. 2056305119836778.

Mattli, Walter and Ngaire Woods, eds. (2009). *The Politics of Global Regulation*. Princeton, NJ: Princeton University Press.

Mayhew, David R. 2004. *Congress: The Electoral Connection*. 2nd ed. New Haven, CT: Yale University Press.

McCarthy, Tom. 2020. *Zuckerberg Says Facebook Won't Be 'arbiters of Truth' after Trump Threat*. The Guardian. URL: http://www.theguardian.com/technology/2020/may/28/zuckerberg-facebook-police-online-speech-trump (visited on 07/15/2021).

McCulloch, Craig. 2019. *Christchurch Call: Tech Companies Overhaul Organisation to Stop Terrorists Online*. Radio New Zealand. URL: https://www.rnz.co.nz/news/political/399468/christchurch-call-tech-companies-overhaul-organisation-to-stop-terrorists-online (visited on 06/08/2020).

McLelland, Mark and Seunghyun Yoo. 2007. "The International Yaoi Boys' Love Fandom and the Regulation of Virtual Child Pornography: The Implications of Current Legislation". *Sexuality Research & Social Policy* 4.1, p. 93.

Medzini, Rotem. 2021. "Enhanced Self-Regulation: The Case of Facebook's Content Governance". *New Media & Society* Early View, pp. 1–25.

Meyer, Trisha. 2017. *The Politics of Online Copyright Enforcement in the EU: Access and Control*. Cham, CH: Palgrave Macmillan.

Microsoft Corporate. 2017. *Facebook, Microsoft, Twitter and YouTube Announce Formation of the Global Internet Forum to Counter Terrorism*. Microsoft On the Issues. URL: https://blogs.microsoft.com/on-the-issues/2017/06/26/facebook-microsoft-twitter-youtube-announce-formation-global-internet-forum-counter-terrorism/ (visited on 07/24/2021).

Mikler, John. 2018. *The Political Power of Global Corporations*. Cambridge, UK: Polity.

Milan, Stefania. 2015. "When Algorithms Shape Collective Action: Social Media and the Dynamics of Cloud Protesting". *Social Media + Society* 1.2, p. 2056305115622481.

Moore, Martin and Damian Tambini, eds. (2018). *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple*. Oxford, UK: Oxford University Press.

Moravcsik, Andrew. 1993. "Preferences and Power in the European Community: A Liberal Intergovernmentalist Approach". *JCMS: Journal of Common Market Studies* 31.4, pp. 473–524.

Morris, Zoë Slote. 2009. "The Truth about Interviewing Elites". *Politics* 29.3, pp. 209–217.

Mueller, Milton, Brenden N Kuerbis, and Christiane Pagé. 2007. "Democratizing Global Communication? Global Civil Society and the Campaign for Communication Rights in the Information Society". *International Journal of Communication* 1.1, pp. 267–296.

Mueller, Milton, John Mathiason, and Hans Klein. 2007. "The Internet and Global Governance: Principles and Norms for a New Regime". *Global Governance: A Review of Multilateralism and International Organizations* 13.2, pp. 237–254.

Müller, Markus M. 2001. "Reconstructing the New Regulatory State in Germany: Telecommunications, Broadcasting and Banking". *German Politics* 10.3, pp. 37–64.

Nash, Victoria. 2019. "Revise and Resubmit? Reviewing the 2019 Online Harms White Paper". *Journal of Media Law* 11.1, pp. 18–27.

Newman, Abraham L. 2017. "Sequencing, Layering, and Feedbacks in Global Regulation". *International Politics and Institutions in Time*. Ed. by Orfeo Fioretos. Oxford, UK: Oxford University Press.

Newman, Abraham L. and David Bach. 2004. "Self-Regulatory Trajectories in the Shadow of Public Power: Resolving Digital Dilemmas in Europe and the United States". *Governance* 17.3, pp. 387–413.

Nieborg, David B. and Thomas Poell. 2018. "The Platformization of Cultural Production: Theorizing the Contingent Cultural Commodity". *New Media & Society* 20.11, pp. 4275–4292.

Nielsen, Rasmus Kleis. 2012. *Ground Wars: Personalized Communication in Political Campaigns*. Princeton, NJ: Princeton University Press.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York, NY: NYU Press.

O'Reilly, Tim. 2007. "What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software". *Communications & Strategies* 65.1, pp. 17–37.

Owen, Taylor. 2015. *Disruptive Power: The Crisis of the State in the Digital Age*. Oxford, UK: Oxford University Press.

Panzic, Stephanie Frances. 2015. "Legislating for E-Manners: Deficiencies and Unintended Consequences of the Harmful Digital Communications Act". *Auckland University Law Review* 21.1, pp. 225–247.

Papaevangelou, Charilaos. 2021. "The Existential Stakes of Platform Governance: A Critical Literature Review". *Open Research Europe* 1.31, pp. 1–11.

Pappalardo, Kylie and Nicolas Suzor. 2018. "The Liability of Australian Online Intermediaries". *The Sydney Law Review* 40.4, pp. 469–498.

Perel, Maayan and Niva Elkin-Koren. 2015. "Accountability in Algorithmic Copyright Enforcement". *Stanford Technology Law Review* 19.3, pp. 473–533.

Pierson, Paul. 2000. "Increasing Returns, Path Dependence, and the Study of Politics". *The American Political Science Review* 94.2, pp. 251–267.

Plantin, Jean-Christophe and Aswin Punathambekar. 2019. "Digital Media Infrastructures: Pipes, Platforms, and Politics". *Media, Culture & Society* 41.2, pp. 163–174.

Plantin, Jean-Christophe and Gabriele de Seta. 2019. "WeChat as Infrastructure: The Techno-Nationalist Shaping of Chinese Digital Platforms". *Chinese Journal of Communication* 12.3, pp. 257–273.

Plantin, Jean-Christophe et al. 2018. "Infrastructure Studies Meet Platform Studies in the Age of Google and Facebook". *New Media & Society* 20.1, pp. 293–310.

Post, Savannah. 2017. "Harmful Digital Communications Act 2015". *New Zealand Women's Law Journal* 1.1, pp. 208–214.

Powell, Alison, Michael Hills, and Victoria Nash. 2010. *Child Protection and Freedom of Expression Online*. Forum Discussion Paper 17. Oxford, UK: Oxford Internet Institute, pp. 1–59.

Powers, Shawn M. and Michael Jablonski. 2015. *The Real Cyber War: The Political Economy of Internet Freedom*. Champaign, IL: University of Illinois Press.

Prime Minister of Australia. 2019. *Tough New Laws to Protect Australians from Live-Streaming of Violent Crimes*. The Office of Hon Scott Morrison MP. URL: https://www.pm.gov.au/media/tough-new-laws-protect-australians-live-streaming-violent-crimes (visited on 07/25/2021).

Puppis, Manuel. 2010. "Media Governance: A New Concept for the Analysis of Media Policy and Regulation". *Communication, Culture and Critique* 3.2, pp. 134–149.

Rath, Christian. 2017. *SPD-Politiker über Facebook-Gesetz: 'Legale Posts wiederherstellen'*. Die Tageszeitung. URL: https://taz.de/!5426108/ (visited on 05/14/2020).

Rauh, Christian. 2019. "EU Politicization and Policy Initiatives of the European Commission: The Case of Consumer Policy". *Journal of European Public Policy* 26.3, pp. 344–365.

Raymond, Mark and Laura DeNardis. 2015. "Multistakeholderism: Anatomy of an Inchoate Global Institution". *International Theory* 7.3, pp. 572–616.

Reinbold, Fabian. 2015. *Hetze auf Facebook: Warum der Hass nicht gelöscht wird*. Der Spiegel. URL: https://www.spiegel.de/netzwelt/web/hetze-auf-facebook-warum-der-hass-nicht-geloescht-wird-a-1051805.html (visited on 07/14/2020).

Renckens, Stefan. 2020. *Private Governance and Public Authority: Regulating Sustainability in a Global Economy*. Cambridge, UK: Cambridge University Press.

— 2020. "The Instrumental Power of Transnational Private Governance: Interest Representation and Lobbying by Private Rule-Makers". *Governance* 33.3, pp. 657–674.

Reuter, Markus. 2017. *: „Allzu restriktiv": OSZE warnt vor Netzwerkdurchsetzungsgesetz*. URL: https://netzpolitik.org/2017/allzu-restriktiv-osze-warnt-vor-netzwerkdurchsetzungsgesetz/ (visited on 07/21/2021).

— 2017. *Bundestagsdebatte: Maas findet sein Hate-Speech-Gesetz gut, alle anderen wollen Änderungen*. netzpolitik.org. URL: https://netzpolitik.org/2017/bundestagsdebatte-maas-findet-sein-hate-speech-gesetz-gut-alle-anderen-wollen-aenderungen/ (visited on 05/14/2020).

— 2017. *Hate-Speech-Gesetz: Geteilte Reaktionen auf den Entwurf des Justizministers*. netzpolitik.org. URL: https://netzpolitik.org/2017/hate-speech-gesetz-geteilte-reaktionen-auf-den-entwurf-des-justizministers/ (visited on 05/11/2020).

— 2017. *Hate-Speech-Gesetz: Neuer Entwurf gefährdet weiterhin die Meinungsfreiheit*. netzpolitik.org. URL: https://netzpolitik.org/2017/hate-speech-gesetz-neuer-entwurf-gefaehrdet-weiterhin-die-meinungsfreiheit/ (visited on 05/14/2020).

— 2017. *Vorsicht Beruhigungspille: Netzwerkdurchsetzungsgesetz geht unverändert in den Bundestag*. netzpolitik.org. URL: https://netzpolitik.org/2017/vorsicht-beruhigungspille-netzwerkdurchsetzungsgesetz-geht-unveraendert-in-den-bundestag/ (visited on 05/14/2020).

Reuters. 2016. *German Minister Tells Facebook to Get Serious About Hate Speech*. Fortune (Reuters Newswire). URL:

https://fortune.com/2016/09/26/heiko-maas-facebook/ (visited on 08/03/2020).

Roberts, Sarah T. 2018. "Digital Detritus: 'Error' and the Logic of Opacity in Social Media Content Moderation". *First Monday* 23.3.

— 2019. *Behind the Screen: Content Moderation in the Shadows of Social Media*. New Haven, CT: Yale University Press.

Rochet, Jean-Charles and Jean Tirole. 2003. "Platform Competition in Two-Sided Markets". *Journal of the European Economic Association* 1.4, pp. 990–1029.

Roger, Charles B. 2020. *The Origins of Informality: Why the Legal Foundations of Global Governance Are Shifting, and Why It Matters*. Oxford, UK: Oxford University Press.

Roose, Kevin. 2021. *In Pulling Trump's Megaphone, Twitter Shows Where Power Now Lies*. The New York Times. URL: https://www.nytimes.com/2021/01/09/technology/trump-twitter-ban.html (visited on 07/15/2021).

Rosenau, James N. and Ernst-Otto Czempiel, eds. (1992). *Governance Without Government: Order and Change in World Politics*. Cambridge, UK: Cambridge University Press. 328 pp.

Rosenblat, Alex. 2018. *Uberland: How Algorithms Are Rewriting the Rules of Work*. Berkeley, CA: University of California Press.

Royal Commission of Inquiry. 2020. *Royal Commission of Inquiry into the Attack on Christchurch Mosques on 15 March 2019*. Wellington, NZ: Department of Internal Affairs, New Zealand Government.

Ruggie, John Gerard. 2007. "Business and Human Rights: The Evolving International Agenda". *American Journal of International Law* 101.4, pp. 819–840.

Rusch, Lina. 2018. *Ein Jahr NetzDG: Grüne und Thinktank fordern Nachbesserungen*. Tagesspiegel: Digital Background. URL: https://background.tagesspiegel.de/ein-jahr-netzdg-gruene-und-thinktank-fordern-nachbesserungen (visited on 05/14/2020).

Sachdeva, Sam. 2019. *A Surprising Ally on Tech Regulation*. Newsroom. URL: https://www.newsroom.co.nz/page/a-surprising-ally-on-greater-tech-regulation (visited on 07/24/2021).

Sarre, Rick. 2017. "Metadata Retention as a Means of Combatting Terrorism and Organised Crime: A Perspective from Australia". *Asian Journal of Criminology* 12.3, pp. 167–179.

Saurwein, Florian. 2011. "Regulatory Choice for Alternative Modes of Regulation: How Context Matters". *Law & Policy* 33.3, pp. 334–366.

Savage, Ashley and Richard Hyde. 2014. "Using Freedom of Information Requests to Facilitate Research". *International Journal of Social Research Methodology* 17.3, pp. 303–317.

Sayers, Anthony M. and Andrew C. Banfield. 2013. "The Evolution of Federalism and Executive Power in Canada and Australia". *Federal Dynamics: Continuity, Change, and the Varieties of Federalism*. Ed. by Arthur Benz and Jörg Broschek. Oxford, UK: Oxford University Press, pp. 185–204.

Schoenebeck, Sarita, Oliver L Haimson, and Lisa Nakamura. 2020. "Drawing from Justice Theories to Support Targets of Online Harassment". *New Media & Society* 23.5, pp. 1278–1300.

Schor, Juliet B. et al. 2020. "Dependence and Precarity in the Platform Economy". *Theory and Society* 49, pp. 833–861.

Schulz, Jacob. 2020. *What's Going on With France's Online Hate Speech Law?* Lawfare. URL: https://www.lawfareblog.com/whats-going-frances-online-hate-speech-law (visited on 07/21/2021).

Schulz, Wolfgang. 2018. *Regulating Intermediaries to Protect Privacy Online: The Case of the German NetzDG*. SSRN Scholarly Paper ID 3216572. Rochester, NY: Social Science Research Network.

— 2019. *Roles and Responsibilities of Information Intermediaries*. Palo Alto, CA: Hoover Institution, Stanford University, pp. 1–28.

Schulz, Wolfgang and Thorsten Held. 2002. *Regulierte Selbstregulierung Als Form Modernen Regierens*. Hamburg, DE: Hans-Bredow-Institut.

Schwemer, Sebastian Felix. 2019. "Trusted Notifiers and the Privatization of Online Enforcement". *Computer Law & Security Review* 35.6, p. 105339.

Seawright, Jason and John Gerring. 2008. "Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options". *Political Research Quarterly* 61.2, pp. 294–308.

Shepherd, Simon. 2019. *The Nation: National Cyber Policy Office's Paul Ash*. Scoop. URL: https://www.scoop.co.nz/stories/PO1909/S00387/the-nation-national-cyber-policy-offices-paul-ash.htm (visited on 10/27/2020).

Sikkink, Kathryn. 1986. "Codes of Conduct for Transnational Corporations: The Case of the WHO/UNICEF Code". *International Organization* 40.4, pp. 815–840.

Silverman, David. 2015. *Interpreting Qualitative Data*. 5th ed. Los Angeles, CA: SAGE.

Singapore Government. 2019. *Protection from Online Falsehoods and Manipulation Regulations 2019*. Singapore Statutes Online. URL: https://sso.agc.gov.sg/SL-Supp/S662-2019/Published/20191001?DocDate=20191001 (visited on 07/21/2021).

Síthigh, Daithí Mac. 2020. "The Road to Responsibilities: New Attitudes towards Internet Intermediaries". *Information & Communications Technology Law* 29.1, pp. 1–21.

Slaughter, Anne-Marie. 2004. *A New World Order*. Princeton, NJ: Princeton University Press.

Smith, Aaron and Monica Anderson. 2018. *Social Media Use in 2018*. Pew Research Center. URL: http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/ (visited on 09/12/2018).

Smith, Brad. 2019. *A Tragedy That Calls for More than Words: The Need for the Tech Sector to Learn and Act after Events in New Zealand*. Microsoft On the Issues. URL: https://blogs.microsoft.com/on-the-issues/2019/03/24/a-tragedy-that-calls-for-more-than-words-the-need-for-the-tech-sector-to-learn-and-act-after-events-in-new-zealand/ (visited on 07/24/2021).

Sonderby, Chris. 2019. *Update on New Zealand*. Facebook Newsroom. URL: https://perma.cc/ZA85-2Y3X (visited on 04/16/2019).

Spindler, Gerald. 2017. "Internet Intermediary Liability Reloaded – The New German Act on Responsibility of Social Networks and Its (In-) Compatibility with European Law". *Journal of Intellectual Property, Information Technology and E-Commerce Law* 8.2, pp. 166–179.

Squirrell, Tim. 2019. "Platform Dialectics: The Relationships between Volunteer Moderators and End Users on Reddit". *New Media & Society*, p. 1461444819834317.

Srivastava, Swati. 2021. "Not Just a Social Network: Facebook as a Private Polity-Maker". 8th European Workshops inInternational Studies (EWIS). Brussels, BE.

Srnicek, Nick. 2016. *Platform Capitalism*. Cambridge, UK: Polity Press.
Stanford World Intermediary Liability Map. 2017. *Act 1068/2006, on the Measures Preventing the Propagation of Child Pornography, December 2006*. URL: https://wilmap.stanford.edu/entries/act-10682006-measures-preventing-propagation-child-pornography-december-2006 (visited on 07/21/2021).
— 2017. *Child Trafficking and Pornography Act, 1998*. URL: https://wilmap.stanford.edu/entries/child-trafficking-and-pornography-act-1998 (visited on 07/21/2021).
— 2017. *Information Technology (Procedure and Safeguards for Blocking for Access of Information by Public) Rules, 2009*. URL: https://wilmap.stanford.edu/entries/information-technology-procedure-and-safeguards-blocking-access-information-public-rules (visited on 07/21/2021).
— 2018. *Computer Crimes Law, June 2009*. URL: https://wilmap.stanford.edu/entries/computer-crimes-law-june-2009 (visited on 07/21/2021).
Statista. 2021. *Most Popular Social Networks Worldwide as of April 2021, Ranked by Number of Active Users*. Statista. URL: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ (visited on 07/15/2021).
Statista Research Department. 2016. *Social Network Memberships in Germany 2016*. Statista. URL: https://www.statista.com/statistics/428657/social-network-memberships-germany/ (visited on 07/21/2021).
Stoker, Gerry. 1998. "Governance as Theory: Five Propositions". *International Social Science Journal* 50.155, pp. 17–28.
Strange, Susan. 1991. "Big Business and the State". *Millennium: Journal of International Studies* 20.2, pp. 245–250.
Streeck, Wolfgang. 2009. *Re-Forming Capitalism: Institutional Change in the German Political Economy*. Oxford, UK: Oxford University Press. 312 pp.
Suzor, Nicolas. 2018. "Digital Constitutionalism: Using the Rule of Law to Evaluate the Legitimacy of Governance by Platforms". *Social Media + Society* 4.3, pp. 1–11.
— 2019. *Lawless: The Secret Rules That Govern Our Digital Lives*. Cambridge, UK: Cambridge University Press.
Tankovska, H. 2021. *Facebook: Average Revenue per User 2011-2020, by Region*. Statista. URL: https://www.statista.com/statistics/251328/facebooks-average-revenue-per-user-by-region/ (visited on 07/21/2021).
— 2021. *Leading Countries Based on Number of Twitter Users as of April 2021*. Statista. URL: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/ (visited on 07/21/2021).
Tansey, Oisín. 2007. "Process Tracing and Elite Interviewing: A Case for Non-Probability Sampling". *PS: Political Science and Politics* 40.4, pp. 765–772.
Taplin, Jonathan. 2017. *Why Is Google Spending Record Sums on Lobbying Washington?* The Guardian. URL: https://www.theguardian.com/technology/2017/jul/30/google-silicon-valley-corporate-lobbying-washington-dc-politics (visited on 09/09/2018).
Tarrow, Sidney. 2005. *The New Transnational Activism*. Cambridge, UK: Cambridge University Press.

Theil, Stefan. 2019. "The Online Harms White Paper: Comparing the UK and German Approaches to Regulation". *Journal of Media Law* 11.1, pp. 41–51.
Thelen, Kathleen and Sven Steinmo. 1992. "Historical Institutionalism in Comparative Politics". *Structuring Politics: Historical Institutionalism in Comparative Analysis*. Ed. by Sven Steinmo, Kathleen Thelen, and Frank Longstreth. Cambridge, UK: Cambridge University Press, pp. 1–32.
Thompson, Peter A. 2011. "Neoliberalism and the Political Economies of Public Television Policy in New Zealand". *Australian Journal of Communication* 38.3, pp. 1–16.
— 2019. "Beware of Geeks Bearing Gifts: Assessing the Regulatory Response to the Christchurch Call". *The Political Economy of Communication* 7.1, pp. 83–104.
Tu, Fangjing. 2016. "WeChat and Civil Society in China". *Communication and the Public* 1.3, pp. 343–350.
Tucker, Joshua A et al. 2017. "From Liberation to Turmoil: Social Media And Democracy". *Journal of Democracy* 28.4, pp. 46–59.
Tufekci, Zeynep. 2017. *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. New Haven, CT: Yale University Press.
Turner, Fred. 2010. *From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism*. Chicago, IL: University of Chicago Press.
Tusikov, Natasha. 2016. *Chokepoints: Global Private Regulation on the Internet*. Oakland, CA: University of California Press.
— 2017. "Transnational Non-State Regulatory Regimes". *Regulatory Theory: Foundations and Applications*. Ed. by Peter Drahos. Canberra, AU: ANU Press, pp. 339–353.
— 2019. "Defunding Hate: PayPal's Regulation of Hate Groups". *Surveillance & Society* 17.1-2, pp. 46–53.
Tworek, Heidi. 2021. "Fighting Hate with Speech Law: Media and German Visions of Democracy". *The Journal of Holocaust Research* 35.2, pp. 106–122.
Tworek, Heidi and Paddy Leerssen. 2019. *An Analysis of Germany's NetzDG Law*. Ditchley Park, UK: Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression, p. 11.
Vaidhyanathan, Siva. 2018. *Antisocial Media: How Facebook Disconnects Us and Undermines Democracy*. Oxford, UK: Oxford University Press.
Van Dijck, José. 2013. *The Culture of Connectivity: A Critical History of Social Media*. Oxford, UK: Oxford University Press.
— 2020. "Seeing the Forest for the Trees: Visualizing Platformization and Its Governance". *New Media & Society* Early View, p. 146144482094029.
Van Dijck, José and Thomas Poell. 2013. "Understanding Social Media Logic". *Media and Communication* 1.1, pp. 2–14.
Van Dijck, José, Thomas Poell, and Martijn de Waal. 2018. *The Platform Society: Public Values in a Connective World*. New York, NY: Oxford University Press. 240 pp.
Van Evera, Stephen. 1997. *Guide to Methods for Students of Political Science*. Ithaca, NY: Cornell University Press.
Van Loo, Rory. 2016. "The Corporation as Courthouse". *Yale Journal on Regulation* 33.2, pp. 547–602.
Vasagar, Jeevan. 2014. *Transcript of Interview with Heiko Maas, German Justice Minister*. Financial Times. URL:

https://www.ft.com/content/766b6116-3cf7-11e4-a2ab-00144feabdc0 (visited on 07/29/2020).

Vernon, Raymond. 1977. *Storm over the Multinationals: The Real Issues*. Cambridge, MA: Harvard University Press.

Vogel, David. 2003. "The Hare and the Tortoise Revisited: The New Politics of Consumer and Environmental Regulation in Europe". *British Journal of Political Science* 33.4, pp. 557–580.

— 2010. "The Private Regulation of Global Corporate Conduct: Achievements and Limitations". *Business & Society* 49.1, pp. 68–87.

Von Notz, Konstantin. 2015. *'Hate Speech': Bundesregierung muss gegen Internet-Hetze vorgehen*. Handelsblatt. URL: https://www.handelsblatt.com/politik/deutschland/hate-speech-bundesregierung-muss-gegen-internet-hetze-vorgehen/12724148.html (visited on 07/21/2021).

Vowles, Jack and Jennifer Curtin, eds. (2020). *A Populist Exception? The 2017 New Zealand General Election*. Canberra, AU: ANU Press.

Wagner, Ben. 2013. "Governing Internet Expression: How Public and Private Regulation Shape Expression Governance". *Journal of Information Technology & Politics* 10.4, pp. 389–403.

— 2016. *Global Free Expression - Governing the Boundaries of Internet Content*. Cham, CH: Springer.

Wagner, Ben et al. 2020. "Regulating Transparency? Facebook, Twitter and the German Network Enforcement Act". Conference on Fairness, Accountability, and Transparency in Machine Learning (FAT*). Barcelona, Spain, p. 11.

Walby, Kevin and Mike Larsen. 2012. "Access to Information and Freedom of Information Requests: Neglected Means of Data Production in the Social Sciences". *Qualitative Inquiry* 18.1, pp. 31–42.

Walby, Kevin and Alex Luscombe. 2017. "Criteria for Quality in Qualitative Research and Use of Freedom of Information Requests in the Social Sciences". *Qualitative Research* 17.5, pp. 537–553.

Weir, Margaret. 1992. "Ideas and the Politics of Bounded Innovation". *Structuring Politics: Historical Institutionalism in Comparative Analysis*. Ed. by Sven Steinmo, Kathleen Thelen, and Frank Longstreth. Cambridge, UK: Cambridge University Press, pp. 188–216.

Weiss, Thomas G. 2000. "Governance, Good Governance and Global Governance: Conceptual and Actual Challenges". *Third World Quarterly* 21.5, pp. 795–814.

Weltevrede, Esther and Erik Borra. 2016. "Platform Affordances and Data Practices: The Value of Dispute on Wikipedia". *Big Data & Society* 3.1, p. 2053951716653418.

Westerwinter, Oliver. 2021. "Transnational Public-Private Governance Initiatives in World Politics: Introducing a New Dataset". *The Review of International Organizations* 16.1, pp. 137–174.

Westerwinter, Oliver, Kenneth W. Abbott, and Thomas J. Biersteker. 2021. "Informal Governance in World Politics". *The Review of International Organizations* 16.1, pp. 1–27.

Witt, Ulrich. 2002. "Germany's "Social Market Economy": Between Social Ethos and Rent Seeking". *The Independent Review* 6.3, pp. 365–375.

Woodcock, Jamie and Mark Graham. 2020. *The Gig Economy: A Critical Introduction*. Cambridge, MA: Polity. 182 pp.

York, Jillian C. 2019. *The Christchurch Call Comes to the UN*. Electronic Frontier Foundation. URL: https://www.eff.org/deeplinks/2019/09/christchurch-call (visited on 04/01/2021).

Ziewitz, Malte. 2016. "Governing Algorithms: Myth, Mess, and Methods". *Science, Technology, & Human Values* 41.1, pp. 3–16.

Zittrain, Jonathan. 2008. *The Future of the Internet and How to Stop It*. New Haven, CT: Yale University Press.

Zürn, Michael. 2018. *A Theory of Global Governance: Authority, Legitimacy, and Contestation*. Oxford, UK: Oxford University Press.