


# INTERNATIONAL POLITICAL ECONOMY AND THE REPUTATION VARIABLE IN IR: THREE ESSAYS

Hayley Pring

Member of

Nuffield College

A DISSERTATION

in

International Relations

Presented to the Faculties of the 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 at the University of Oxford

Feb 2025

Word Count: 73,391

Supervisor of Dissertation

Dr. Alex Kuo, Professor of Government, Department of Politics and International Relations, University of Oxford

Dissertation Committee

Professor Krystof Pelc, Professor of International Relations, University of Oxford
Professor Don Casler, Professor of Political Science, University of Illinois at Urbana-Champaign

INTERNATIONAL POLITICAL ECONOMY AND THE REPUTATION VARIABLE IN IR: THREE ESSAYS

COPYRIGHT

Feb 2025

Word Count: 73,391

Hayley Pring

# ACKNOWLEDGEMENT

I am indebted to a very long list of people for making this dissertation and my graduate studies possible.

To begin with, I have to start with my doctoral supervisor, Alex Kuo. Alex stepped in to supervise me in the last 18 months of my DPhil. What was initially an informal arrangement quickly became a full supervision. Alex's vast knowledge and methodological expertise were invaluable throughout the project. Alex was incredibly kind, helpful, and always willing to give me time when I needed his help or to review my work (often with not much notice). Alex pushed me to develop my ideas, helped me pivot when one project fell apart, and provided me with invaluable guidance. I am extremely grateful to Alex for his help these last two years, and feel very lucky to have had his guidance. Second, I have to thank my original supervisor, Karolina Milewicz, who took over as my supervisor during the last year of my MPhil. Karolina gave me a lot of latitude to morph my dissertation away from its original project origins in economic statecraft, into a new direction that reflected my long-standing interests in organizational behavior and political psychology in IR. Though my projects shifted to cover diverse topics linked by their focus on reputation, Karolina was always encouraging. Despite leaving Oxford, Karolina stayed in touch and has organized workshops to help her students.

Beyond my direct supervisors, I have to thank Duncan Snidal, who has read and provided feedback on my work for several years now. Duncan has been more than generous with the amount of time and care he puts into his role as a supervisor (and informal supervisor in my case!). From workshops, seminars, lunch meetings, and volunteering his time as part of my confirmation panel, Duncan and his legendary work in IR has inspired me greatly along the way. From a devastatingly funny comment on not burning an early draft of my paper on climate change because it would be bad for the environment and instead to just recycle it, to a kind and helpful intervention with a

project that inadvertently took a wrong direction in the last year of my DPhil, Duncan's humour, understanding and deep and unrivaled knowledge of the discipline of IR is a true asset to Oxford and Nuffield College, and I feel very lucky that I have managed to cross paths with him.

At Nuffield College, I need to thank a huge host of people that turned an academic institution into a thriving and welcoming community. To Eleni, the college tutor, I have to thank her for her constant help and calmness when I would inevitably reach out in panic. Eleni is the foundation of the community at Nuffield College. I have to thank my boss and friend, Professor Pepper Culpepper, who has provided me with a great deal of advice and wisdom over the years, and a whole lot of kindness. I will also thank the successive and committed college supervisors I've been lucky to have, from Andy Eggers who inspired my keen interest in causal inference techniques and interrogating our methods, to Janina Dill, whose feedback and assistance provided me with some much needed guidance. The Nuffield College IR lunch seminars were a highlight of my time there.

Overwhelmingly I need to thank Nuffield for its generosity. It is no mistake that graduate studies in the UK are an expensive pathway, and because of this, it's often a very unequal and unrepresentative cohort. However, Nuffield College's generosity is the reason that there is more diversity and inclusion present in the graduate community at Oxford. Through the Nuffield Scholarship, I was able to continue my DPhil studies. I am grateful for the assistance of the college during the pandemic. I feel extremely lucky for the housing (and I must thank the consistent help of Claire Bunce here!), offices space, and travel grants, that enabled me to carry out my research. The free lunches (there is such a thing as a free lunch at Nuffield) led me to meeting a whole host of individuals I would never have had the chance to speak with. The staff at Nuffield cannot be overlooked, from the funny and kind Porters - Tony, to the staff that runs the college - I could never forget Olivier Goddett, an extremely funny and immensely kind presence at Nuffield that I looked forward to seeing everyday.

Related to this are the granting institutions, of which I have several to thank! I am a first-generation high-school graduate, and every institution that has granted me a scholarship has changed my life. I want to thank Nuffield College, the Governor-Phillip Scholarship, the DPIR at Oxford, the James-Fairfax Oxford Australia granting body, and the Gowrie Foundation for their incredible generosity that I am very, very indebted to. Thank you to the Brett-Schneider Fund for enabling me to travel to Cornell University and learn from some excellent scholars in the field. I will endeavour to give back to scholarships foundations in the future, who do so much for social mobility, the future of research, and changing people's lives.

Previous version of my project benefited from questions, criticism, and suggestions at Oxford's DPIR seminars and workshops, Nuffield College IPE workshop, workshops hosted by Karolina and Duncan, and the University of Bayreuth, Gottenberg, Cornell University, the American Political Science Association, the Political Economy of International Organizations, and the International Political Economy Symposium. I have to thank the Computational Social Science workshops, the European Consortium, and various Oxford workshops for the extremely helpful feedback. There are many individuals whose feedback and comments helped improve this dissertation.

In the DPIR community, my friends and fellow-supervisees provided not only critical feedback and a make-shift supervision for me, but endless laughs and great conversations. Claas Mertens and Sam Seitz, two exceptional scholars who will make great waves in the world, and whose advice was absolutely invaluable. A special place in this acknowledgements must go to my very close friend, Viviana Baraybar Hidalgo. Viviana and I went through the highest of highs and the lowest of lows together, she is hands-down one of the funniest, smartest, and most resilient people I know. I have a text message saved from her in response to the price of a flight I needed to take home, with the response "you don't really need 2 kidneys?" I look forward to hearing about future coups in Peru. I have to also thank many friends for their help (either academically or otherwise over the many years), Laure Bokobza, Julia de Romémont , Charlotte Grace, Matty Burrage, Sam Seitz, Claas

Mertens, Giuseppe Spatafora, Gerda Raissar.

My final thanks goes to my family and close friends. My parents, Olive and Andrew Pring, and my sister Bronwyn May Pring. If dissertation topics are at all mirrors to who we are as people, it merits mentioning that my family endeavoured to give me every opportunity in the world, supported me at every turn, and always gave me the option of coming home. I know we cannot choose our family but I feel so fortunate to have mine. My family were my support network, thousands of miles away in Western Australia, and I felt that everyday for five years. To my very close friends who saw me through emotional upheavals, rejections, celebrations, and uncertainty, I have to thank my best friends Mariah Mathew and Isadora (Dodie) Bogle. Mariah for her twenty plus years of friendship, her enormous heart, kindness, generosity, unrelenting support, and ability to make me laugh while crying. Dodie for her support, distractions, wonderful conversations, and always being somebody I can turn to. Perhaps I have overused the term 'lucky' in this dedication, but there is no other way of saying it, I am richer for the both of them, and I would not have made it through this without them.

# Table of Contents

**Title Page** i

**Acknowledgement** iii

**Abstract** vii

**Table of Contents** viii

1 **Introduction** 1

1.1 Crisis and Reputation ..... 2

1.2 Current Debates on Reputation and IR ..... 4

1.3 Reputation as a Group-Based Theory ..... 8

1.4 Heroic vs incremental reputational costs ..... 10

1.5 Firms, Reputation and Cooperation Problems in IR ..... 12

1.6 Concluding Remarks ..... 14

1.7 References ..... 16

2 **Literature Review** 20

2.1 Reputation in International Relations ..... 20

2.1.1 Defining Reputation ..... 20

2.1.2 Self-Image vs Reputation ..... 21

2.1.3 Signaling Reputation and the necessity of follow-through ..... 24

2.2 History Dependence of Reputation ..... 26

2.3 Crisis and Reputational Types ..... 32

2.3.1 Crisis ..... 33

2.4 Situational vs Dispositional Reputations ..... 35
2.5 Psychology and the role of Group Behaviour in Reputation ..... 40
2.6 Mechanisms of Reputation ..... 44
2.6.1 Cascades ..... 45
2.6.2 Externalities and Network-Effects ..... 52
2.7 Reputation as a Dynamic Group Effect ..... 53
2.8 Dueling Audiences ..... 61
2.9 Firms and Leaders: Non-State Actors with Reputations ..... 60
2.10 Incremental versus Heroic Sanctions ..... 64
2.11 References ..... 67

3 **Paper One: Reputation Between Disciplines:**
**A Group-Based Theory of Firm Reputation in IR**
81
3.1 Introduction ..... 81
3.2 Reputation between Disciplines ..... 89
3.2.1 Defining Reputation ..... 89
3.3 Signaling and attributes ..... 91
3.3.1 Economics ..... 91
3.3.2 IR ..... 93
3.4 Reputational Types and The Nature of the Audience ..... 98
3.5 History Dependence of Reputation ..... 101
3.6 A Group-Based Theory of Reputation ..... 107
3.7 Reputation as a Dynamic Group Effect ..... 111
3.8 Pathways of Reputation: Cascades, Externalities, and Early-Movers
114
3.8.1 Cascades ..... 115

3.8.2 Early-Movers ..... 121
3.8.3 Dueling Audiences ..... 124
3.9 Conclusion ..... 127
3.10 References ..... 130

4 **Paper Two: Who Gives a Damn About a Bad Reputation? The Corporate Exodus and the Russian-Ukrainian War** 146

4.1 Introduction ..... 148
4.2 Sanctions and Firm Reactions ..... 156
4.3 A Theory of Reputation and Intra-Industry Diffusion ..... 158
4.4 Multi-Stakeholder Initiatives: Reputation-Insulation ..... 171
4.4.1 UNGC: Background & Membership ..... 172
4.4.2 UNGP: Background & Membership ..... 175
4.5 Data ..... 177
4.5.1 Reputation: Measuring Diffusion Effects ..... 180
4.5.2 Social Media Statements ..... 183
4.6 Research Design ..... 186
4.7 Empirical Analysis ..... 190
4.8 Firm Statements: Analysis ..... 206
4.9 Conclusion ..... 223
4.10 References ..... 230
4.11 Appendix ..... 243

5 **Paper Three: Who Needs To Be Seen To Be Green? How Reputational Pressure Affects Responses To Climate Change** 300

5.1 Introduction ..... 302

5.2 The Carbon Disclosure Project: Background ..... 308

5.3 Variation in Firm Behavior on Climate Change ..... 310

5.4 A Reputational Theory of Firm-Level Climate Responsiveness ..... 313

5.4.1 Natural Disasters ..... 318

5.4.2 Reputational Externalities ..... 320

5.4.3 Reputational Cascades ..... 322

5.5 Research Design ..... 326

5.5.1 Case Study: Australia ..... 326

5.5.2 Data ..... 327

5.5.3 Explanatory variables ..... 327

5.5.4 Controls ..... 330

5.6 Results ..... 331

5.6.1 Natural Disasters ..... 336

5.6 Conclusion ..... 345

5.7 References ..... 348

5.8 Appendix ..... 367

6 **Discussion and Conclusions** 391

6.1 The Reputation Variable in IR ..... 391

6.2 Reputation for Non-State Actors in International Political Economy 402

6.3 The Necessity of Crisis ..... 404

6.4 Domains of Reputation and its effectiveness ..... 405

6.4.1 Reputation and incremental costs ..... 406
6.4.2 Reputational change and permanency ..... 409
6.5 Reputation as a Relational Concept ..... 4511
6.6 On Dueling Audiences and Conflicting Reputations ..... 416
6.7 References ..... 425

# ABSTRACT

## INTERNATIONAL POLITICAL ECONOMY AND THE REPUTATION VARIABLE IN IR: THREE ESSAYS

Hayley Pring

Reputation is a fundamental idea in International Relations (IR), used to explain why states go to war, or why they might cooperate despite the absence of formal enforcement mechanisms. Surprisingly, however, we know little about the varieties of reputation and how it shapes agents' preferences and behavior. In particular, how new types of agents on the international stage experience reputation. To this end, this dissertation is concerned with three key questions: 1) How does the group-effects of reputation affect firm behavior 2) What effect do different types of crisis across domains of IR have on generating reputational costs and how actors respond? and 3) How do agents strategically manage multiple audiences and conflicting reputations among those audiences? Through these three questions, this dissertation seeks to interrogate three key aspects that underpin conventional wisdom on the role of reputation in IR: the agent, the audience, and the role of crisis. Though states have occupied the predominant actor position in the literature on reputation in IR, this paper centralizes the role of firms and political leaders.

Reputation is generally discussed in IR as an intangible asset of the state. Reputation affects other agents' attitudes towards that state, and hence the focus is on the role of reputation in influencing a state and their capacity to trade, invest, form alliances, comply with international law, or their ability to deter aggression. A great deal of research has sought to mediate whether a state's 'reputation' results from situational traits, that is to say, derived form the strategic environment in which different actors come to the same conclusion about another agent's reputation, or dispositional traits; that is, characteristics or attributes of actors themselves. When IR scholars discuss reputation they tend to discuss individual leaders' reputations as temporary or distinct from a state's reputation. For instance, Dafoe and Caughey (2016) discuss leaders' concern for honor and

how it can lead to a preference for fostering reputations for resolve; Guisinger and Smith (2002) emphasize individual leader honesty and its importance in signaling credibility in diplomacy, and Kertzer (2016) asks why some leaders display resolve during confrontations and crises, while others choose not to. Parallel to this is work by Guisinger and Smith (2002), Chiozza and Choi (2003), and Wolford (2007) who emphasize the relationship between tenure and institutions that mold those reputations. In this body of work, reputations of states are divorced from reputations of leaders. Similarly, the current literature on reputation often assumes state preferences to be a homogeneous block – leaving out the burgeoning growth of non-state actors and how they may factor into a domestic governments concern over reputation, or whether the reputational concerns of firms trumps that of the state in certain domains or international events. In this way, the relationship between the reputation of the individual or group of elites, and the reputation of the collective state itself, is unclear. It's an idea entertained in passing by Jervis (2017) who explores how the reputation of the statesman is used as a focal point for the public to form perceptions of the reputation of the nation state, but Jervis never goes beyond the singular statesman. This dissertation seeks to explore the role of firms in the reputation literature and how concern for reputation can influence firm behavior.

This leads to the secondary contribution of these papers; deconstructing the observers of reputation. Extant work on reputation has assumed an amorphous single audience (or group) places costs on an actor for failing to either live up to their obligations, show resolve, comply with international obligations, or violate their reputational type in general. Audience costs are measured either through the penalty imposed on the actor, or in an ex-ante setting, by examining how the actor behaves to reduces the costs they anticipate they will incur from damaging their reputation among this single audience. This dissertation challenges this concept by introducing multiple 'conflicting' audiences across all three papers in different contexts, and addresses the question: how do agents respond to situations where they hold different reputations among different audiences? And, any decision involves a trade-off, where an agent is rewarded by one group and punished by another? When actors make choices that result in reputational penalties being imposed upon them, it's not necessarily an indication that the actor is insensitive to these costs – rather it may indicate that they are making

choices to satisfy one audience, while mitigating the damage to their reputation with a competing audience. These situations generally arise in periods of crisis or high issue salience, where actions are difficult to conceal, and therefore hiding from one group and appeasing another becomes challenging.

Crisis plays a critical role in the literature on reputation. As in the original formulation of Fearon (1994)'s work on reputation and war, crisis heightens the consequences and increases the prominence of an actor's decisions and behavior by raising the stakes of audience costs and the added drama of time preferences. His predecessor on signaling, Schelling (1960), makes note of the importance of crisis for reputation, where observers are more likely to notice strong signals by using heuristic short-cuts to make judgments. As (ibid., p. 11) states, "And the reason why such evidence might be forthcoming in a crisis is that the motive to provide it – the greater urgency of reaching an understanding or an agreement that depends on it – may be enhanced in such an emergency". But the dominant crises that typifies studies of reputation are conflict signaling and war. This dissertation explores two areas of crises across the domain of political economy; economic statecraft and geopolitical crisis and climate-change crisis. As the nature of the crisis changes across these domains, so too does the type of reputation that matters, and how actors must behave to appease audiences.

The nature of the crisis, the actor, and their relevant audience effects the type of reputation that matters. Unlike previous studies that have honed in on the idea of reputations for *resolve*, this dissertation explores reputational types related to how audiences perceive 'good' and 'bad' types during these crisis; firms that abide by human rights standards during war, firms that are trying to improve and not contribute to adverse climate-change induced disasters, and leaders that attempt to maintain particular anti-establishment reputations while maintaining close ties to those economic elites.

The three papers that make up this dissertation proceed as follows:

The first paper of this dissertation explores the evolution of reputation in two sibling disciplines: international relations and economics and management. I argue that while both literatures had similar starting points, reputation evolved distinctively between the disciplines as IR focused on states, the structural constraints of the anarchical international system, and the resulting need to focus on resolve. In contrast, the institutional features of firms and markets allowed for a more flexible, variegated, and multi-faceted - branching off into an array of sub-disciplines, while for IR, reputation is largely dominated by security studies. Since firms are structurally distinct from states, and the international system is in many ways different to markets, integrating both perspectives is an effective way to evaluate how concern for firm reputation in international relations. In this first paper I focus on all three questions in the dissertation, but I offer an overarching framework for how reputation is considered. I argue that, while historical behavior is still important for reputational formation, if observer knowledge is low as a function of a less-salient reputation, then observer's will use heuristics and easily accessible information to make judgments on type: an actor's peers. By introducing a group-based theory of reputation, I explain the conditions under which concern for reputation can induce firms to cooperate, self-regulate, or adjust their behavior. A group-based theory of reputation also has implications for how reputation diffuses. I introduce the notion of cascades, early-movers, and dueling audiences as pathways through which reputation operates.

The second paper in this dissertation explores the role of reputation during geopolitical crisis for firms. I address three central ideas of reputation. First, reputations are not formed in vacuums; rather reputations are formed by the behavior and decisions of a firm's competitors and group. Second, firms balance conflicting reputations among dueling audiences. And third, firms utilize different strategies to appeal to these reputational types. Following Russia's invasion of Ukraine in 2022, some multinational corporations (MNCs) went beyond compliance with sanctions, and issued statements on voluntarily withdrawing business operations in Russia, while other firms paused operations or continued 'business as usual'. Why did some MNCs issue statements to stay and others

to leave? I make two key arguments. First, I argue that firms that issued statements to withdraw from Russia did so to minimize the reputational costs of remaining amid online backlash. Rather than concern for preserving a history of socially-conscious behavior, early withdrawal statements by peer firms created a 'reputation stealing' effect, and forced within-industry firms to withdraw within close intervals of their competitors. I test this argument using cox proportional hazard models, and OLS with fixed-effect models using data collected from the Yale Russian Withdrawal dataset, and novel data on multiple time-points of company statements across social media platforms. Second, I argue that firms that remained did so to preserve politically-neutral reputations among a rival audience, Russia, in order to secure their safe return to the Russian market in the future. These firms employed specific strategies to mitigate reputational costs. Using word-embedding techniques and topic modeling on a corpus of company statements, I show that remaining firms are less critical of Russia, less likely to issue social media statements, and more likely to leverage the legitimacy of membership to UN-adjacent multi-stakeholder initiatives. I find that remaining firms are also more likely to attempt to join these public-private partnerships post-invasion, suggesting bad actors learn how to leverage legitimacy from IO-adjacent organizations. This study provides insights into how reputation affects corporate responses to geopolitical shocks.

The final paper in this dissertation explores reputation during climate crises. This paper addresses reputation at the firm level, focusing on how variation in reputational pressure effects firm incentives to disclose information on carbon mitigation. The paper builds on the findings of paper one and paper two, by again reinforcing the idea that the effect of reputation is not distributed homogeneously among st agents. I develop a four-pronged theory of reputational pressure to explain variation in firm disclosures. First, firms face different reputational pressure based on their primary consumer audience. Business-to-consumer facing firms face greater reputational costs of failing to act on climate, and therefore avoid exposure. Second, 'climate-conscious' reputations matter when consumer audiences are 'primed' to value them. I argue that climate disasters induce consumer awareness, and firms respond by signaling to audiences their 'climate-conscious' reputation. Third, reputations have cascade effects. Firms are more likely to disclose if a large early-mover within

their sector also discloses. Fourth, there is a trade-off for firms when signaling action on climate; attention brings greater scrutiny, which may damage the firms reputation more than keeping quiet. Following poor climate evaluations, firms become much less likely to disclose in the future. Using Australia as a case study, I employ data on firm disclosure and climate-mitigating action from the Carbon Disclosure Project (CDP) between 2010-2021. In addition to natural disaster data from EMDAT, I proxy for the material effects of natural disasters using insurance claims, and use a staggered difference-in-difference estimator to compare high and low years following climate disasters. The empirical tests show reputational pressure can both induce and paralyses firm action on climate.

In sum, rather than a single problem, this dissertation addresses significant gaps in the international relations literature on reputation by offering a systematic theoretical and empirical re-examination of reputation of reputation by focusing on firms and integrating multi-disciplinary insights. It offers an overarching theory of reputation as a group based concept. In doing so, it contributes to several strands of research on reputation. This dissertation shows that the role of reputation in influencing behavior in IR is contingent on the relationship between the agent and the competing audiences they rely on for survival. Furthermore, I show that reputation is not 'fixed' but instead bench-marked to standards set by peers, and that reputation should be understood as a group-based concept, not just a historically-determined one. The findings suggest that concern for reputation can in some instances encourage more compliant, cooperative, or consistent behavior from actors – but often it can also lead firms to either switch to more covert strategies to avoid antagonizing one audience when they attempt to appease another, pressures actors into making quick decisions for short-term gain with long-term costs, and nudges them into choosing actions that have greater purchase among one audience, even if those actions have very little substance and may be less effective. The findings in all three papers suggest that the visibility of the actor to the audience is more important for reputation in influencing behavior than being the most powerful agent in a given crisis.

# INTRODUCTION & LITERATURE REVIEW

In 2021, an enormous multi-national company at the heart of rising domestic and international concerns over data privacy, election interference, terrorism, and misinformation campaigns changed its iconic parent company name from 'Facebook' to 'Meta'. The corporate re-branding served the function of signaling the company's intentions to move beyond social media, but it also represented an attempt to distance itself from the growing toxicity of its reputation among consumers, stakeholders, regulators, and governments. The move to abandon a household name was bold, but not without precedent. Google restructured itself under the parent name 'Alphabet' in 2015 so that the company would no longer be associated with its risky and controversial endeavors in other areas, including their expansion into the field of artificial intelligence ventures in Natural Language Processing. Philip Morris, the tobacco behemoth, suffered decades of reputational damage over the global publicity and condemnation of the hazardous health effects of smoking. The company finally attempted to repair its reputation by re-branding into the Altria Group in 2001. For corporations, like states, reputation is an intangible organizational asset (C. J. Fombrun 2005; Dierickx and Cool 1989; Barney 1991). Reputation is a resource that is both precious and inimitable – it cannot be traded nor substituted, but it can lend the object of that reputation an enduring competitive advantage over their peers. Unlike in International Relations, where scholarly work continues to debate the power of reputation in shaping state behavior – or whether reputation matters at all (Lebow 1981; Orme 1992; Hopf 1994; Mercer 2010) - micro-economics, finance, and the literature on business organization are unanimous in their regard for the critical role of reputation in both making or breaking a company, as well as motivating firm behavior (Milgrom and Roberts 1982; Milgrom and Roberts 1986; Shapiro 1983; Spence 1974; Weigelt and Camerer 1988), with work on 'signaling' and reputation a much belabored field (C. Fombrun, Van Riel, et al. 2003; Dierickx and Cool 1989; Ravasi et al. 2018).

Scholarly attention in IR has recently and increasingly shifted towards the role of Multi-National Corporations (MNCs) in influencing a range of outcomes in IPE, including trade policy (Kim and

Osgood 2019; Kim 2017; Baccini, Pinto, and Weymouth 2017; J. B. Jensen, Quinn, and Weymouth 2017; Osgood 2017), shaping the formation of preferences around trade (Lee and Liou 2022), affecting cooperation over climate governance (Kennard 2020; Colgan and Hinthorn 2023; Colgan, Green, and Hale 2021; Cory, Lerner, and Osgood 2021; Lerner and Osgood 2023; Rowan 2021), observance over human rights regimes (Thrall 2021), compliance with international agreements (N. M. Jensen and Malesky 2018), the expansion of judicial institutions and judicial power (Kalyanpur and Newman 2019), the effectiveness of sanctions and economic statecraft more broadly (Bapat and Kwon 2015), and the fragility of the Liberal International Order (Flaherty and Rogowski 2021). However, even among this burgeoning body of research, only peripheral attention has been paid to the role of reputation in influencing firm behavior in IR.

In focusing on firms as the major actors, this dissertation pivots from the traditional focus on states, and brings into question the role of heterogeneous audiences with conflicting preferences, how these audiences can generate different and sometimes conflicting reputational types, the influence of group-dynamics on reputation, and how reputational concerns can affect behavior across different issue areas in IPE.

## Crisis and Reputation

At the beginning of his term Joe Biden invoked the language of overcoming a "stain on our reputation" and declared the need for the US to "salvage our reputation".[1] The comments were made in reference to the outgoing Trump administration, and the withdrawal of the US, both formally and informally, from the international order after 2016. Concern for reputation – and indeed how actions can signal reputational types – is littered throughout the discourse of world leaders. In a 2013 speech to at the National Conference on Organizational Work, Xi Jinping proclaimed, "Officials need to 'cherish their reputation'. This depends on what kind of 'reputation' they are cherishing.

[1]Comments made in: Joseph R. Biden, Jr. 23 January 2020. "Why America Must Lead Again: Rescuing U.S. Foreign Policy After Trump" Foreign Affairs. Remarks by President Biden at Signing of Executive Orders Advancing His Priority to Modernize Our Immigration System. 02 February 2021.

Is it a 'reputation' which will be applauded by people with ulterior motives, or is it a reputation for acting in the interests of the party and the people? A party member should only cherish the latter reputation, and it would be calamitous if he were bent on gaining the former!".[2] The idea that Jinping was invoking was that officials would face conflicting incentives to cater to different audiences with different reputations. The notion of reputational types is also evident in the remarks of David Cameron in 2010, "Britain's reputation, our reputation in the world. It is not just about might. It's about doing what is right".[3] Signaling particular reputations, to particular audiences, motivated Germany's response to the refugee crisis of 2015, Angela Merkel explained, "It is our reputation that is at stake, our image in the world in view of civil war and terrorism".[4] Observably, it is clear that certain reputations are appealed to in order to justify or explain behavior in international relations across a range of different issues areas.

While some research is motivated by a paucity of attention to an idea, or the problematic neglect of a phenomenon, this dissertation is not that. Reputation itself is one of the most frequently invoked independent variables in the study of International Relations. It is a ubiquitous ingredient in IR, used to explain everything from why states go to war, why they win wars, and why they cooperate over common interests in the absence of any formal enforcement mechanism and incentives to defect. The exhaustive use, but often misuse, of reputation proliferates political actors' words in both motivations and justifications, and underlines many of the theories of world politics and the grand theories of Realism, Liberalism, and Constructivism. The aim of this dissertation is not to pretend to reintroduce a saturated idea, but rather to unpack a nebulous, unmeasured and as it is currently used – unhelpful – concept. The intangibility of observing reputation, and measuring reputation, is part of its misleading application across the study of global politics.

[2]Xi Jinping. 28 June 2013. The governance of China: Train and Select Good Officials. Quishi: CPC Central Committee Bimonthly

[3]BBC News. 6 October 2010.David Cameron's speech in full

[4]Angela Merkel on refugee policy: A European solution is in Germany's interests. 29 February 2016. Bundesregierung.

At its broadest level, this dissertation is motivated by three interlocked puzzles through the prism as both agents with reputations: 1) How does the group-effects of reputation affect firm behavior 2) What effect does different types of crisis have on generating reputational costs and how do actors respond? and 3) How do agents strategically manage multiple audiences and maintain conflicting reputational types? IR has sought to explain behaviors and outcomes in international politics by pointing to variation in a very specific area of reputation – that is, resolve – but this mission has limited the development of the study of reputation in explaining other types of variation. In short, by restricting the study of 'reputation' in IR to just states and the issue area to just conflict, we have very little understanding of the differential effects of reputation on actors, and which issue areas or types of crisis might make actors more sensitive to reputational costs. The consistent focus on resolve as the key reputational type has limited the study of reputation in world politics, and with it, a theory of reputation itself.

This dissertation advances a number of ideas that complicates the simple notion of reputation currently used in IR. The contributions can be broken down into the type of actor who develops a reputation (firms), issue area (geopolitical crisis, and climate change), the role of conflicting reputations and dueling audiences, and critically – that reputation is a group-based concept that develops in reaction to how audiences observe relative behavior to an actor's peers.

## Current Debates on Reputation and IR

The first part of this dissertation explores the evolution of reputation in two sibling disciplines: international relations and economics and management. I argue that while both literatures had similar starting points, reputation evolved distinctively between the disciplines as IR focused on states, the structural constraints of the anarchical international system, and the resulting need to focus on resolve. In contrast, the institutional features of firms and markets allowed for a more flexible, variegated, and multi-faceted - branching off into an array of sub-disciplines, while for IR, reputation is largely dominated by security studies. Since firms are structurally distinct from states,

and the international system is in many ways different to markets, integrating both perspectives is an effective way to evaluate how concern for firm reputation in international relations. In this first paper I focus on all three questions in the dissertation, but I offer an overarching framework for how reputation is considered. I argue that, while historical behavior is still important for reputational formation, if observer knowledge is low as a function of a less-salient reputation, then observer's will use heuristics and easily accessible information to make judgments on type: an actor's peers. By introducing a group-based theory of reputation, I explain the conditions under which concern for reputation can induce firms to cooperate, self-regulate, or adjust their behavior. A group-based theory of reputation also has implications for how reputation diffuses. I introduce the notion of cascades, early-movers, and dueling audiences as pathways through which reputation operates.

Multinational corporations themselves are increasingly at the center-stage of world politics on transnational issues. I explore this phenomenon in paper two. Following the invasion of Ukraine in February 2022, Russia was hit with a swathe of almost historic economic penalties by Western European countries, the United States and their allies. The imposition of these punishing sanctions were not limited to governments. Corporations began acting early in the conflict – the energy sector departing swiftly – at first buying time, then suspending operations, and finally, in some cases, withdrawing entirely from the state. Under pressure from consumers, stakeholders, and investors, the corporate exodus from Russia represented the power of concern for reputation. Take for instance the statement issued by British Petroleum in February 2022 in explaining their shareholder exit from Rosneft, directly citing "the impact on our reputation". Or the statement issued by the chief executive of one of the largest dairy giants operating in Russia, Danone, "in the end our reputation is about our behavior". Several Japanese automobile manufacturers commented on the risk to their reputation several days before suspending operations and signaling withdrawal. These private sanctions were critical in crippling the Russian economy in the absence of any reasonable form of military escalation from NATO-affiliated countries. Firms have become pivotal agents in IR, but reputation – and risk to reputation – from their audiences often shape the behavior of these agents by nudging them into behaving in non-rational ways.

The second and third papers of this dissertation extend the logic of reputation to the level of the firm, as a major and growing agent in International Relations. The second paper in this dissertation explores the role of reputation during geopolitical crisis for firms. I address three central ideas of reputation. First, reputations are not formed in vacuums; rather reputations are formed by the cascade effects of peer behavior. Second, firms balance conflicting reputations among dueling audiences. And third, firms utilize different strategies to appeal to these reputational types. Following Russia's invasion of Ukraine in 2022, some multinational corporations (MNCs) went beyond compliance with sanctions, and issued statements announcing their intention to voluntarily withdraw from Russia, while other firms paused operations or continued 'business as usual'. Why did some MNCs issue statements to stay and others to leave? I make two key arguments. First, I argue that firms that issued statements to withdraw from Russia did so to minimize the reputational costs of remaining amid online backlash. Rather than concern for preserving a history of socially-conscious behavior, early withdrawals by peer firms created a 'reputation stealing' effect, and forced within-industry firms to withdraw within close intervals of their competitors. I test this argument using fixed-effect models and cox proportional hazard models with data collected from the Yale Russian Withdrawal dataset, as well as novel data on multiple time-points of company statements across social media platforms. Second, I argue that firms that remained did so to preserve politically-neutral reputations among a rival audience, Russia, in order to secure their safe return to the Russian market in the future. These firms employed specific strategies to mitigate reputational costs. Using word-embedding techniques and topic modeling on a corpus of company statements, I show that remaining firms are less critical of Russia, less likely to issue social media statements, and more likely to leverage the legitimacy of membership to UN-adjacent multi-stakeholder initiatives. I find that remaining firms are also more likely to attempt to join these public-private partnerships post-invasion, suggesting bad actors learn how to leverage legitimacy from IO-adjacent organizations. This study provides insights into how reputation affects corporate responses to geopolitical shocks.

The third paper again addresses reputation at the firm level, focusing on how variation in reputational pressure effects firm incentives to disclose information on carbon mitigation. The paper builds on the findings of paper one and paper two, by again reinforcing the idea that the effect of reputation is not distributed homogeneously among agents. Agents experience varied reputational pressure depending on their audience type - and concern for reputation can induce sub-optimal behavior among agents.

In this third paper I focus on climate governance and firms. Private actors are crucial to the most salient problem in IR; international cooperation on climate change. Domestic and international laws have been slow and ineffective at regulating firms in response to climate change. For this reason, voluntary self-regulation and private regulatory bodies have become increasingly important. Despite claims that firms oppose costly carbon mitigation strategies, previous research has shown variation in the extent to which firms signal their performance on climate change. What explains this variation? I develop a four-pronged theory of reputational pressure to answer this question. First, firms face different reputational pressure based on their primary consumer audience. Business-to-consumer facing firms face greater reputational costs of failing to act on climate, and therefore avoid exposure. Second, 'climate-conscious' reputations matter when consumers are 'primed' to value them. I argue that climate disasters induce consumer awareness, and firms respond by signaling to audiences their climate-conscious reputation. Third, reputations have cascade effects. Firms are more likely to move to disclose following large early-movers within their sector. Fourth, there is a trade-off for firms signaling action on climate; attention brings greater scrutiny, which may damage the firms reputation more than keeping quiet. Following poor climate evaluations, firms become much less likely to disclose in the future. Using Australia as a case study, I employ data on firm disclosure and climate-mitigating action from the Carbon Disclosure Project between 2010-2021. In addition to natural disaster data from EMDAT, I proxy for the material effects of natural disasters using insurance claims, and use a difference-in-difference estimator to compare high and low years following climate disasters. The empirical tests show reputational pressure can both induce and paralyze firm action on climate. Perversely, I find that concern for reputation means that transparency can

deter disclosure.

In all three papers, I endeavor to address how concern for reputation can motivate firms to either cooperate or self-regulate. I argue that reputation exerts a strong influence over agent behavior across cooperation problems in political economy. All three papers that make up this dissertation address the relationship between two sides; the agent on one side, and the audiences on the other, with whom they must develop, maintain, or repair reputations in response to changing attitudes or preferences. This interaction between audiences and agents when it comes to reputation has been a missing link in the literature. Although IR has dealt extensively with reputation as a causal (but largely untested) variable to explain conflict initiation and cooperation under conditions of uncertainty, it has ignored the behavioral elements of reputation. Reputations are not fixed, they change and adapt to audience expectations, and they are far from one dimensional. This dissertation argues that 'reputation' can operate through several mechanisms to induce behavioral change. By integrating insights from behavioral science, micro-economics and theories of organizational behavior, this dissertation advances the depth of reputational studies in IR. Two of the three papers deals with a separate area of International Political Economy (IPE) – geopolitics, economic statecraft, and climate change. While these topics are very different, the breadth of case studies serves to demonstrate the context-dependent nature of reputations' effects on actors, and the profound consequences non-tangible ideas like reputation can have on world politics. I adopt a variety of causal inference techniques and endeavor to formally and quantitatively test the role of reputation. While capturing 'reputation' as a variable is enormously difficult, its presence has proved so consequential in explaining agent's behaviors.

### Reputation as a group-based theory

An argument that runs across all three papers is that reputations are not developed in a vacuum, nor are they fixed at all points in time, and more often than not they are contingent on the relative benchmark set by an agent's peers, or by the expectations they set in advance through their appeal

to certain audiences. The functionality of reputation is at the heart of the literature on reputation; whereby a government's reputation becomes an important asset they foster to convince others to enter into agreements with them. Where the literature looks exclusively at states as the main actors, it makes sense that the contextual conditions, such as international regimes, are the focus, since international regimes help governments assess reputations by providing standards of behavior against which performances could be measured, by linking these standards to specific issues, and by providing forums, often through international organizations, in which these evaluations can be made. However, this dissertation pushes back against the idea that international regimes are what create benchmarks. While it is true that these institutions initially set the standards of behavior, it is peer behavior that reinforces whether or not an individual, or a collective, will be punished for following or disavowing that behavior.

Despite the prolific work of Keohane on the role of reputation in international organizations and its role in encouraging cooperation, even his own findings suggest that reputations might be unstable and 'one set of rules' might not be the standard that everybody in a system (or regime) is held to (Keohane 1984). Ironically, Keohane came to the realization that international regimes provided 'benchmarks for behavior' after reading Elizabeth Colson's account on how stateless societies reached consensus on the character of individuals (ibid.). However, what Colson says is quite different to one set of rules for all individuals to follow; it is that "onlookers apply the standards of performance in particular roles in making overall judgment about the total person; this in turn allows them to predict future behavior. Judgment is an ongoing process through which consensus is finally reached" (Colson 2017, p. 53). Rather than one set of behaviors for all, standards are bench-marked to A) the role you are occupying, and B) the process of judgment is adjusted over time, as perhaps your role changes as does the standards of behavior around that role. The second paper shows this shifting reputational costs dynamic as peer behavior influences the social expectations of the 'acceptable' benchmark among consumers. The third paper shows a similar effect of peer-to-peer behavior through cascade effects.

Heroic vs incremental reputational costs

One of the other well-worn foundational areas of theories of reputation in IR is the punishment or penalty imposed on the actor for violating their reputational type'. Actor A commits action X which violates their reputational type, and an audience imposes a singular 'cost' on the actor. In most theories of resolve this involves one large sanctioning effort - perhaps the domestic audience votes out the leader who failed to show resolve after signaling their commitment to fight, or perhaps the international audience refuses to cooperate with the actor that failed to comply with an international agreement or treaty. The sanction is all or nothing - if the leader is voted out of office then there was a cost to them violating their reputational type, but if the leader remains in office, there is no penalty. Or perhaps in the case of the company that suffered reputational costs from remaining in Russia, or failing to signal action on climate change, did they suffer some large decreases in profit? Or, did employees and/or consumers stage a large boycott? These might be classified as a 'heroic' sanctioning effect that illustrates the effects of reputational costs on the actor.

These examples are excellent at showing the effect of reputation on conditioning actors behavior - either guiding their action, or explaining why an actor might be sanctioned - however they also obfuscate instances where actors are guided by smaller and less absolute costs, that accumulate over time. Costs are incurred by political leaders, not just through elections held at various time points, but also through shifts in public support between elections. Similarly, costs are felt by companies, not just through cuts in profit, but through slander or misinformation that may affect their brand image - foretelling some future effect on profit even if it's not realized immediately. What I mean by this is that there are 'incremental sanctioning' effects (Coleman 1994), whereby an actor known to have a particular reputational type, commits action X that violates this type, and certain members of the audience impose smaller retaliatory costs on the actor, however these costs are additive and contagious and ultimately have a larger reputational cost on the actor. This idea of 'incremental sanctioning' in the sociological literature demonstrates how norms are enforced, not through one large sanctioning effort, but by smaller efforts over time (ibid.).

This brings me to a secondary idea on network and cascade effects. I show across the papers in this dissertation that there is a network effect created by peers who are the bearers of the reputation. However, there is also a similar network effect among the audience of those imposing costs. An individual, or a handful of individuals, who constitute audience members, impose a penalty on an actor they perceive to be violating their reputation. That penalty may be as simple as snubbing or ignoring the transgressor, but in the case of the consumer it could be through the boycotting of products, or the degradation of the company's status among consumers. In these instances the actions of one or a few individuals (or activists) who are audience members to the actor, can have a ripple effect on other audience members. Not just through the sharing of information about the actor that could discredit their reputation, but witnessing one audience member impose costs could lead another audience member to impose similar costs out of fear that not doing so could be viewed by their peers as inadvertently endorsing that actors behavior, or the audience member might reap their own social benefits from imposing costs by their peers. Without the idea of 'incremental sanctioning' effects, this would be lost by IR scholars who are simply looking for one large penalty at one point in time - rather than the accumulation of smaller penalties from consumers, or voters, or party-supporters, over time. All three papers in this dissertation attempt to shine a light on this incremental sanctioning effect by examining how consumers and voters might penalize an actor slowly over time, which eventually leads the company or the individual to change course to protect their reputation from further damage.

The relationships among individual audience members conditions how they impose costs on the offender. To alter an example offered by Coleman (1994, p. 281), say one member of a club sanctions the leader of the club for their poor behavior - perhaps the power imbalance might mean that the leader can silence that member, or alienate that member from the other club members. This cost is high enough that it might deter the individual from making the effort to criticize the leader. In some ways, the audience members have their own reputations to protect among the audience they

belong to, that might be threatened if they go against the norms among that audience, and impose social costs on the leader. However, say over time the behavior of that club leader is such that more than one member of the audience begins to step forward, it lowers the costs for other audience members to express disapproval as well, creating enough momentum overtime for reputational costs to be imposed on the leader, which may lead to tangible outcomes. This is a pretty simple analogy that is meant to demonstrate that for some actor-audience relationships, the relationships among audience members can condition when, and the extent to which ,they can impose costs on the actor.

Re-examining what a reputational cost is in IR broadens the field into exploring more than just 'audience costs' as one singular penalty imposed by voters, consumers, or members of treaties, law, and international agreements. The power imbalances in IR, whether that be the material resource differences between alike states, or the relative power of some leaders over others, or the size and resources of firms, means that the imposition of a single collective cost on an actor is usually beyond the means of the audience member(s). Just as is the case in imposing sanctions in the IR literature, an audience member can be easily deterred from imposing reputational costs on an actor if the audience member themselves will be punished - either directly by their peers and the social consequences of going against the group, or by the actor themselves should the individual be vocal or important enough to garner attention from the actor. Papers two and three explore the idea of incremental sanctions guiding behavior over time, rather than one large reputational cost.

### Firms, Reputation and Cooperation Problems in IR

An additional contribution in this dissertation relates to whether reputation can serve as an enforcement mechanism for cooperation of firms in IR. In the neoliberal-institutionalism literature, reputation plays a central role akin to its role in deterrence; unlike realists, neoliberal institutionalists believe that changes in preferences over strategies to reach particular outcomes can result in mutual benefits. This change is brought about by decreasing informational asymmetries related to a given situation, motivations, and deductions about what the other side will do in the future

(Jervis 1999). Under conditions of anarchy, where enforcement mechanisms are absent or simply ineffective, reputation is perhaps the central disciplinary mechanism that audiences can use to force agents to behave, to force them to comply, and to punish them for non-compliance with specific normative standards. For transnational issues, like climate change, or the modernizing tactic of fighting geopolitical conflicts through economic statecraft, or reaching and maintaining agreements on trade, corporations are crucial to cooperative outcomes. With the existing patch-work of regulations across countries and opaque domestic lobbying, finding mechanisms that might improve cooperative outcomes is critical.

Paper two and three paint a complicated picture when the firm is the actor who must either maintain or defend their reputation. Paper two suggests that reputation can discipline firms into behaving, but only if the issue area is salient enough, and this effect declines over time. The drawbacks, however, suggest that firms strategically manage their reputations through 'blue-washing' strategies, and often leverage international organization membership to deflect from non-compliant behavior. The third paper shows a less optimistic picture; when the action itself is voluntary (such as disclosure of emissions), transparency can actually deter compliance. While in paper two, increased transparency over which firms issued statements to leave or remaining in Russia led to a greater likelihood of corporate retreats from Russia, in the case of paper three, when high-emitting firms face greater exposure from consumers through climate shocks, they're less likely to voluntarily disclose their emissions.

All three papers challenge the argument that the behavior of firms, or agents more generally, is determined situationally. That is, the approach that suggests that, during a crisis all actor's react to the same uniform audience costs. The reaction is not determined by differences in who the actors are; their disposition or the unique characteristics of the actor. The famous example in IR is, of course, Arnold Wolfer's analogy on house fires. "Imagine a number of individuals, varying widely in their predispositions, who find themselves inside a house on fire. It would be perfectly realistic

to expect that these individuals, with rare exceptions, would feel compelled to run toward the exits... Surely, therefore, for an explanation for the rush for the exits, there is no need to analyze the individual decisions that produced it" (Wolfers 1962). But to extend Wolfer's analogy in the context of this dissertation; why did some people exit first? Did the location of the individuals in the house affect how long it took them to exit, or the way in which they exited? Did all individuals exit through the front-door? The baseline finding across all three papers is that the decision to leave the house, the timing, and the mode of exit is determined by the peers within the house. When the first person leaves through the front-door, the rest are likely to follow. These papers argue that the type of actor does matter, because the type of actor and their relationship with an audience, or *audiences*, shapes both their sensitivity to reputational costs, as well as their response.

| Paper Comparison of Contributions | | |
|---|---|---|
| IPE Area | Geopolitics & Sanctions | Climate Change |
| Actors | Firms, IOs | Firms, IOs |
| Reputation Type | Human Rights | Environmental |
| Audience Type | Citizens, State | Citizens, Businesses |
| Reputation Mechanism | Consumer pressure | Consumer pressure |
| Reputation Effect | Cascade | Cascade |
| Crisis | War | Natural Disaster |

## Concluding Remarks

In sum, this dissertation does not resolve a singular problem in IR, rather it addresses significant gaps in the IR literature by offering a systematic theoretical and empirical re-examination of firm reputation in world politics. In doing so, it contributes to several strands of research on reputation, including the relationship between an actor's reputation and the audience that observes and makes judgments on this reputation. It explores different reputational types and how and why they develop

in reaction to crisis; how the nature of the audience affects the signals actors send; and finally that reputational concerns can both encourage and discourage cooperation and self-regulation of multinational firms. It also contributes to broader debates on whether firms will impose private sanctions on bad actors in times of war, and why some firms will self-regulate on climate change while others hide to avoid public scrutiny.

# Bibliography


Baccini, Leonardo, Pablo M Pinto, and Stephen Weymouth (2017). "The distributional consequences of preferential trade liberalization: Firm-level evidence." In: *International Organization* 71.2, pp. 373–395.

Bapat, Navin A and Bo Ram Kwon (2015). "When are sanctions effective? A bargaining and enforcement framework." In: *International Organization* 69.1, pp. 131–162.

Barney, Jay (1991). "Firm resources and sustained competitive advantage." In: *Journal of management* 17.1, pp. 99–120.

Chiozza, Giacomo and Ajin Choi (2003). "Guess who did what: Political leaders and the management of territorial disputes, 1950-1990." In: *Journal of Conflict Resolution* 47.3, pp. 251–278.

Coleman, James S (1994). *Foundations of social theory*. Harvard university press.

Colgan, Jeff D, Jessica F Green, and Thomas N Hale (2021). "Asset revaluation and the existential politics of climate change." In: *International Organization* 75.2, pp. 586–610.

Colgan, Jeff D and Miriam Hinthorn (2023). "International Energy Politics in an Age of Climate Change." In: *Annual Review of Political Science* 26, pp. 79–96.

Colson, Elizabeth (2017). *Tradition and Contract: The Problem of Social Order*. Routledge.

Cory, Jared, Michael Lerner, and Iain Osgood (2021). "Supply chain linkages and the extended carbon coalition." In: *American Journal of Political Science* 65.1, pp. 69–87.

Dafoe, Allan and Devin Caughey (2016). "Honor and war: Southern US presidents and the effects of concern for reputation." In: *World politics* 68.2, pp. 341–381.

Dierickx, Ingemar and Karel Cool (1989). "Asset stock accumulation and sustainability of competitive advantage." In: *Management science* 35.12, pp. 1504–1511.

Fearon, James D (1994). "Domestic political audiences and the escalation of international disputes." In: *American political science review* 88.3, pp. 577–592.

Flaherty, Thomas M and Ronald Rogowski (2021). "Rising inequality as a threat to the liberal international order." In: *International Organization* 75.2, pp. 495–523.

Fombrun, Charles, Cees Van Riel, et al. (2003). "The reputational landscape." In: *Revealing the Corporation: Perspectives on identity, image, reputation, corporate branding, and corporate-level marketing*, pp. 223–33.

Fombrun, Charles J (2005). "Corporate reputations as economic assets." In: *The Blackwell handbook of strategic management*, pp. 285–308.

Guisinger, Alexandra and Alastair Smith (2002). "Honest threats: The interaction of reputation and political institutions in international crises." In: *Journal of Conflict Resolution* 46.2, pp. 175–200.

Hopf, Ted (1994). *Peripheral visions: Deterrence theory and American foreign policy in the third world, 1965-1990*. University of Michigan Press.

Jensen, J Bradford, Dennis P Quinn, and Stephen Weymouth (2017). "Winners and losers in international trade: The effects on US presidential voting." In: *International Organization* 71.3, pp. 423–457.

Jensen, Nathan M and Edmund J Malesky (2018). "Nonstate actors and compliance with international agreements: An empirical analysis of the OECD anti-bribery convention." In: *International Organization* 72.1, pp. 33–69.

Jervis, Robert (1999). "Realism, neoliberalism, and cooperation: understanding the debate." In: *International security* 24.1, pp. 42–63.

– (2017). *How statesmen think: The psychology of international politics*. Princeton University Press.

Kalyanpur, Nikhil and Abraham L Newman (2019). "Mobilizing market power: Jurisdictional expansion as economic statecraft." In: *International Organization* 73.1, pp. 1–34.

Kennard, Amanda (2020). "The enemy of my enemy: when firms support climate change regulation." In: *International Organization* 74.2, pp. 187–221.

Keohane, Robert O (1984). *After hegemony*. Vol. 54. Princeton: Princeton university press.

Kertzer, Joshua (2016). *Resolve in international politics*. Vol. 2. Princeton University Press.

Kim, In Song (2017). "Political cleavages within industry: Firm-level lobbying for trade liberalization." In: *American Political Science Review* 111.1, pp. 1–20.

Kim, In Song and Iain Osgood (2019). "Firms in trade and trade politics." In: *Annual Review of Political Science* 22, pp. 399–417.

Lebow, Richard Ned (1981). *The Nature of International Crisis*.

Lee, Haillie Na-Kyung and Yu-Ming Liou (2022). "Where you work is where you stand: A firm-based framework for understanding trade opinion." In: *International Organization* 76.3, pp. 713–740.

Lerner, Michael and Iain Osgood (2023). "Across the Boards: Explaining Firm Support for Climate Policy." In: *British Journal of Political Science* 53.3, pp. 934–957.

Mercer, Jonathan (2010). *Reputation and international politics*. Cornell University Press.

Milgrom, Paul and John Roberts (1982). "Predation, reputation, and entry deterrence." In: *Journal of economic theory* 27.2, pp. 280–312.

– (1986). "Price and advertising signals of product quality." In: *Journal of political economy* 94.4, pp. 796–821.

Orme, John D (1992). *Deterrence, reputation and Cold-War cycles*. Springer.

Osgood, Iain (2017). "The breakdown of industrial opposition to trade: Firms, product variety, and reciprocal liberalization." In: *World Politics* 69.1, pp. 184–231.

Ravasi, Davide et al. (2018). "The formation of organizational reputation." In: *Academy of Management Annals* 12.2, pp. 574–599.

Rowan, Sam S (2021). "Does institutional proliferation undermine cooperation? Theory and evidence from climate change." In: *International Studies Quarterly* 65.2, pp. 461–475.

Schelling, Thomas C (1960). "Reciprocal measures for arms stabilization." In: *Daedalus* 89.4, pp. 892–914.

Shapiro, Carl (1983). "Premiums for high quality products as returns to reputations." In: *The quarterly journal of economics* 98.4, pp. 659–679.

Spence, Michael (1974). "Competitive and optimal responses to signals: An analysis of efficiency and distribution." In: *Journal of Economic theory* 7.3, pp. 296–332.

Thrall, Calvin (2021). "Public-private governance initiatives and corporate responses to stakeholder complaints." In: *International Organization* 75.3, pp. 803–836.

Weigelt, Keith and Colin Camerer (1988). "Reputation and corporate strategy: A review of recent theory and applications." In: *Strategic management journal* 9.5, pp. 443–454.

Wolfers, Arnold (1962). *Discord and collaboration: essays on international politics*. Baltimore: Johns Hopkins Press.

Wolford, Scott (2007). "The turnover trap: New leaders, reputation, and international conflict." In: *American Journal of Political Science* 51.4, pp. 772–788.

# Literature Review

## 1 Reputation in International Relations

### 1.1 Defining Reputation

All studies of a nebulous concept like *reputation* in international relations must begin with a clarification of definition. I'll firstly outline the most dominant definition of reputation before I propose additional amendments. Mercer (2010, p. 6) defines it as "Reputation is a judgment of someone's character (or disposition) that is then used to predict or explain future behavior". This tracks with similar definitions, such as Dafoe, who defines it as, "Reputation refers to beliefs about an actor's persistent characteristics or behavioral tendencies based on the latter's past behavior" (Dafoe, Renshon, and P. Huth 2014). Renshon, Yarhi-Milo, and J. D. Kertzer (2023) uses a more socially-conscious definition of reputation as "socially shared beliefs about an actor's behavioral tendencies", and Yarhi-Milo (2018) opts for a vaguer definition that divorces reputation from the necessity of historical behavior, here reputation is, "the belief that others hold about a particular actor". The variation in whether reputation requires observation of past actions is clear in these three definitions.

All of these definitions suggest important criteria for reputation that deserves some interrogation. The first assumption is that all dispositional traits

or characteristics point to a particular type of reputation. As I detail below, the indices, signals and behavior used to characterize a reputational type may be interpreted differently by observers - or suggestive of multiple types of reputations. The assumption also infers the vantage point of all observers is the same; different vantage points could produce different perceptions of the agent's type or different emphasis on certain indices of behavior, and therefore vantage point could condition observer's judgment of an actor's reputation. Second, the assumption underpinning both of these definitions is that reputation is history dependent. I will come back to this point further below, but it is important to keep in mind that a history-dependent reputation assumes that all observers have i) no informational asymmetries, where observers know the history of the agent or that this history was transparent and easy to recall, and ii) that the history of the agent is cumulative, and that every observer places the same weight on each action of the agent and therefore the agent's behavior at a single time-point is somehow less informative than their behavior at past time-points. Lastly, iii) the implication is that observers, not bearers, of reputation make these judgments independently from each other.

## 1.2 Self-Image vs Reputation

I will consider this last point first; the relationship between self-image and reputation. IR scholars usually distinguished between self-image and reputation, largely because reputations are not properties that an agent, usually the state, is able to own and independently decide upon. Indeed it is because reputation is not a property that it is ultimately a group-based concept in that the observer of the reputation is the one that makes the judgments of the actor and it is this process that the reputation itself develops. Though, this premise

is not maintained consistently throughout different types of literature. For instance, research on reputation in microeconomics and labor organizations posits a strong assumption that an actor's behavior and actions operate in isolation from others. Reputation is thought to be some sort of asset that the bearer of the reputation has direct control of in building up and influencing (Wilson 1985; Hurwicz 1946). Even in Keohane and Nye (2001, p. 289)'s iconic *Power and Interdependence*, they remark that reputations constitutes one of a state's most important assets, referring to the remarks of Kindleberger, "In economics bygones are bygones, but in politics they are working capital". In this conceptualization of reputation, the bearer is at least aware of their reputation, and are actively cultivating it. Whether reputation is truly independent from self-image is therefore a theoretical issue.[1]

Some of the logical steps that are taken in order to distinguish reputation from self-image are not just an issue of semantics, but are confusing and problematic. For instance, Mercer (2010, p. 27) contends that, "If A acts out of concern for A's reputation, then it is self-perceptions that matter, not reputation...". Mercer goes onto use the example of British foreign policy in the 19th century, "England's reputation was damaged after it appeared to break its promise to assist Denmark against Prussia and Austria in 1864; Britain believed its reputation was damaged and that this belief probably led it to be more cautious.... Britain's self-image took a beating, not its reputation".[2] Awareness of how a reputation is changing - where Britain's consciousness of how its behavior was being perceived led it to make different decisions - cannot purely be self-image motivated. Concern for reputation means the agent manipulates, signals, or uses different indices to try and influence how observers perceive the

[1] For Snyder and Diesing (2015) a statesman's perception of their reputation among observers is itself uncertain, even if they believe they have previously demonstrated resolve.

[2] Mercer's example is a reference from Jervis (1989) page 89.

actors reputational type. The actor's belief that they are type A for instance, may lead them to want others to view them as type A, but they could presumably still imagine themselves to be type A, while they're simultaneously aware that others view them as type B. Being self-aware of reputation does not mean that the agent is acting out of self-image concerns. A large theoretical gap in the literature is acknowledging that concern for reputation is a motive that influences behavior independently (or perhaps overlapping with) concern for self-image. As Jervis highlights, "...Actors who know that they are being observed and want to influence others have powerful incentives to project images that will lead observers to draw the desired impression whether or not it is accurate" (Jervis 2017, p. 109). Being aware of how an agent's behavior might influence one of their aggregate reputational types is exactly how it is possible for observers to de-couple reputations from particular cases.[3] Again, Jervis suggests this, "Of course, actors may not have Signaling reputations overall, but Signaling reputations in particular areas. Thus I might be believed to be ready to carry out my promises but not my threats; to send valid Signals about my capabilities but not about my intentions; to be willing and able to keep my word in the economic area but not on military issues" (ibid., p. 119). If we maintained Mercer's position on the distinction between self-image and reputation then decoupling is a conceptually impossible task, we could not identify when actors behave or make decisions out of concern for reputation (or at least disentangling when they acted out of concern for reputation versus self-image would be messy), and actor's would not be able to signal a reputation independently from the knowledge of their reputational type, or the type of reputation they are intending to acquire, build-up, or even break. All three of my papers suggest that actors are conscious of their reputation, and actions are led by either a

[3]For instance the state argues following a crisis that failure to show resolve (or backing down) was a compromise, as the example is given of Khrushchev in the case of Cuba in 29162 by Snyder and Diesing (2015).

desire to capitalize on a situation to build up a reputation, fear of developing of a poor reputation, or awareness of competing reputational types and how to cultivate oppositional reputations among dueling audiences. In this way, my dissertation contributes to breaking down the hard distinction in the literature between reputation and self-image.

## 1.3 Signaling Reputation and the necessity of Follow-through

The second task when setting out the parameters of reputation is to distinguish it from signaling. As discussed, an actor's reputation involves general beliefs that are held by observers about the actor's type. By actor's type I mean that observers make inferences based on behavior about an actor that are informative and can help classify how and which actors may make decisions now or in the future. The concept has its origin in the work of Jervis (1989) *Logic of Images*. Behaviors that discriminate between possibilities help observers make decisions about how to engage with, react to, or perceive a threat. As highlighted in the *Logic of Images*, the process of making inferences is complicated by the ability of actor's who are incentivize to project an image of themselves that may lead observer's into making inferences about the actor's type that may not be true. The deception that Jervis discusses about states in world politics, is in someway more true of non-state actors who may be responding to multiple audiences. I depart, however from Jervis's definition of signaling reputation as being the point in which an actor follows through on what they said they would do. This definition fits more with reputations for resolve, where resolve refers to maintaining a policy or stance despite temptations or motives to back down. Instead, I would suggest that actions actors take that reaffirm the 'type' they attempt to project is a more befitting outline of signaling reputation. What

Jervis is getting at when he outlines reputational signaling, is credibility. This is reinforced when he states, "Thus I might be believed to be ready to carry out my promises but not my threats; to send valid Signals about my capabilities but not about my intentions; to be willing and able to keep my word in the economic area but not on military issues."(Jervis 2017, p. 119). Jervis is suggesting actors can send signals on many different reputational types, however the carrying through, keeping their word, and validity are all references to credibility, which, as discussed above, is different to a reputational type. After all, actors can develop reputations of being fickle or non-credible in one area, but not anther.

Asymmetric information would allow for an actor to project one image of themselves, successfully, without necessarily having to follow through. Here, the temporal element of behavior and observation conditions the necessity of follow-through. I offer evidence to this argument in my second paper on corporate behavior during the Ukraine crisis, where it is clear that a firm can successfully claim to have departed Russia following the countries invasion of Ukraine, without necessarily following through on this action. The firm may even reverse position and go back on their word. The asymmetry of information for the consumer means that they may not be aware of the follow-through, especially as the salience of the situation declines over time with media coverage, and attention is diverted elsewhere. My dissertation makes a key contribution to definitions and conceptualization of reputation as necessarily involving follow-through by attempting to show that asymmetrical information and salience of information can render the necessity of follow-through redundant. Similarly, situations that trigger audience awareness heighten the costs of not following-through, as my third paper shows on firm disclosures on climate behavior following natural

disaster shocks that prime audiences.

## 2 History Dependence of Reputation

The dominant assumption in IR is that reputation is a history-dependent concept. For both the literature on reputations for resolve and cooperation, the premise that underpins these theories of behavior is that commitments are interdependent. This might be across issue area, regions, and actors (Weisiger and Yarhi-Milo 2015).[4] In some versions of this theory, a state will observe how state A behaves towards state B in a given dispute, and use this information as a template to anticipate behavior in a different dispute, in a separate arena. For instance, Jervis (2017, p. 85) claims that, "Past behavior can be used as an index to the state's characteristics and how it will behave in the future in a way that does not depend on signals". Empirical examples of this include Treisman (2004) who illustrates the logic of Kreps et al. (1982) 'chain store paradox' model in IR by showing how states' invest in reputations for appeasement which seemingly slowed the disintegration of the British Empire. Treisman (2004) suggests that under conditions of resource scarcity, investing in reputations for appeasement in one area can help a state face multiple challengers at once. Sechser (2010) has pushed against the idea that reputation building does not matter in times of crisis, and uses the 1939 Russo-Finnish crisis to argue that adversaries anticipate reputational costs of backing down, and offer side payments to offset these costs, thereby avoiding costly wars. Military crises are partially dependent since these histories of behavior communicate cost tolerance. Orme (1992) argues that Soviet leaders used past behavior of the US to respond to

[4]Reputation being conceptualized as a historically-built phenomenon stems from rationalist understandings of IR. While rationalist definitions classify reputation as "the degree to which an actor reliably upholds its commitments, based on a record of that actor's past behavior". In contrast, constructivists differ by suggesting "reputation is the generally shared opinion of a referent's character or nature, based on a wide range of information, associations and social cues" (Sharman 2007, p. 20).

conflict to judge the country as possessing a poor reputation for resolve. And Press (2005) finds little evidence that German or US policy-makers were able to infer behavior of each other based on past records of resolve. Press argues temporal dynamics mean governments only pay attention to situational attributes over dispositional attributes during crisis. His argument is that, contrary to conventional wisdom, credibility is not related to a state's "history of keeping or breaking commitments" (ibid., p. 143), but observers believing "threats and promises that are backed by sufficient power and serve clear interests" (ibid., pp. 8–9). A big issue for this argument is that the audience of observers are a smaller pool of experienced decision-makers with information on the state's past behavior; the idea that all states - or indeed all actors - have a blank canvas in every crisis is difficult to maintain. For both Press (ibid.) and Mercer (2010), the traditional realist variables of current, material interest and power are what explains outcomes in crisis.

Yet there is by no means a consensus in the literature. Empirical work examining whether historical behavior did build reputation that influenced policymakers has found limited evidence (Hopf 1994; Richard 1981; Snyder and Diesing 2015). Wolford (2007) argued that new leaders face incentives to build reputations for resolve in the shadow of leadership turnover, while antagonists are not deterred by a state's past resolve, since new incumbents may possess differing levels of resolve. The result is that past behavior is irrelevant, and adversaries will seek to test new incumbents. Challenging the idea of T. Schelling (1960) on indiscriminate reputation building across time and space are Snyder and Diesing (2015), Jervis (1989), and P. K. Huth (1997). For P. K. Huth (1988), historical behavior of a state only impacted current behavior to the extent that it was in reference to the same adversary with whom it shared a

history of conflict (Danilovic, Clare, and Tucker 2020). Similarly, in order for historical behavior to matter for reputations for resolve as a deterrent, other scholars have found there must be a common variable (or perhaps, a common focal point). For instance, Danilovic (2002) found that when a trajectory of conflict occurs in the same region, past behavior of one member of a dyad can influence how the other member of the dyad responds.

There are two issues with the current debate about the relevance of history-dependence in reputation formation. The first is that scholars that are seeking to argue against game-theoretic models that past behavior does not impact current behavior do not have a suitable argument for explaining how reputations form in the absence of historical record. The assumption then, is that reputation in general neither matters as an independent variable, nor does it seemingly matter as an outcome variable; if reputation does not matter, actors in world politics should not be motivated to create, maintain, or avoid specific types of reputations. For some of these scholars, it might be the case that their argument only holds for reputations for resolve, not reputations in other lower-stake arenas, such as economic relations. However, given the overwhelming focus on resolve, there has not been enough work to establish whether this is the case. The second issue with this debate is that explanations for why history should not matter in reputation building are premised on variables that only really apply to states, not non-state actors. For example, Tang (2005) argues that the anarchic features of the international system prevent reputations from forming. Additionally, imperfect and asymmetric informational environments mean that observers and audience members could form different reputational perspectives on actors, or alternatively the audience members could take cues from each other on how their peers view an actor's reputation. The bearers of reputation belong

to groups and this can influence their behavior, but observers of reputation also belong to groups that can influence their judgments.

In light of this tension in the IR literature, the organizational and labor economics literature has some profound and interesting insights on this. Despite some claims from IR scholars that "game theorists and economists...these scholars are uninterested in how reputations form. Instead, they assume the conditions that must obtain for a reputation to form and then use the concept (as an independent variable) to solve problems" (Mercer 2010, p. 3), I would argue that more headway has been made in this literature than has been achieved in IR, mostly because these scholars are willing to consider more actors and different types of reputations. Tirole (1996) has been responsible for the most compelling and groundbreaking work on how reputations form in his work on *A Theory of Collective Reputations*. In this work, Tirole also advances a formal model of the interdependence between collective and individual reputations, but also that reputation must necessarily be history-dependent because all generations of actors have overlap. In this model, reputation is history-dependent, however the referent's behavior could encompass a wide array of potential behaviors, not just an ability to maintain commitments. Tirole does outline the conditions under which reputation can be reset; i.e. amnesty, or revolution - where there is a complete overhaul and therefore no overlap in actors with reputations for corruption or cheating. One lasting implication of Tirole's model on generational reptuations that seems to translate well to the literature on state reputations and IR is that reputations are sticky, and it is difficult to shake being tarred as a dishonest type (Mearsheimer 2011; Sartori 2002).

My dissertation seeks to fill this gap by suggesting that history-dependence

is less relevant in these situations, because audiences use heuristic short-cuts with the most accessible information to form ideas on an actor's reputation. For instance, evaluating firms by their nearest-neighbors, or industry-type. I bridge the gap between these two literatures by focusing on non-state actors (firms, and leaders), and considering how a larger universe of competing agents, and a multiplicity of audiences changes the logic of horizontal (historical) versus vertical (nearest neighbor) assessments of reputation. Thus, rather than an either-or argument over history versus situation (disposition versus situation in other formulations), it might be the case that conditions over number of actors in a situation, salience of past events, the nature of the audience(s) and their ability to perceive signals, and the time-horizon of the actors involved, all factor into the extent to which history matters for reputation formation. While my argument here is that, for firms, nearest-neighbor behavior in a situation and observer comparisons within types matters more for formation of reputation, this is not incompatible with the idea that for states (and international organizations), or situations involving one or a small pool of actors, reputations are history dependent.

I want to lastly touch on studies using conjoint experiments and survey experiments that vary past behavior to assess reputations for resolve or credible commitment seem to assume that the experimental setting of prompting the participant to recall an instance of past behavior where the actor either did or did not follow through could be translated easily to a real situation (Tomz 2007; Renshon, Dafoe, and P. Huth 2018; Casler, Ribar, and Yarhi-Milo 2023; J. D. Kertzer, Renshon, and Yarhi-Milo 2021; Tingley and Walter 2011). Under conditions of hyper-aware participants, where the past incident was highly salient among the general public, or the past incident does not date back too far,

then it does seem reasonable to test the effect of past behavior on observer's assessment of current behavior. The role of past behavior (vertical comparisons) in influencing how observers assess reputation likely does hold for prominent states or certain leaders, where either the time horizon is not too long ago in the case of the latter, or the event is highly salient in the case of the former. However, when we introduce a large universe of actors (firms), operating in different sectors, where audiences (consumers) may not be exclusively interacting or residing within that single actor - or aware of past behavior - under these conditions situational elements and horizontal comparisons between alike actors are more likely reference points for observers to make judgments over reputation. It is in this context of many competing actors at once, such as is the case with firms in two of these three papers in this dissertation, that the ability for individuals to recall specific incidents of past behavior for each firm, where the past event was either highly salient or occurred in short succession, seems extremely unlikely. The deficit in the literature on how the conditions under which observers form ideas of reputation vary by actor and reputational type has largely been the result of the continued focus on states as the units of observation, or the treatment of leaders as developing the same reputations.

With these considerations in mind, I offer a competing definition of reputation in this dissertation, that reputation is "beliefs that a group shares about an actor's behavior in a setting". This definition is not explicit about the group-based and horizontal nature of reputation, however this dissertation suggests that the context and domain of behavior shape which type of reputation matters, and while shared beliefs among the observers of a reputation is a necessary criteria of reputation, these beliefs might differ between groups.

# 3 Crisis and Reputational Types

The preoccupation in the IR debates over whether reputations for resolve are "one of the few things worth fighting over" overlooks the role of different types of reputation in non-conflict issue areas, and how - whether theorists believe or not a reputation is worth fighting for - actors of all different shapes and sizes will do whatever they can to protect, maintain, and create specific types of reputations to fit their agenda. Extant work on different reputational types in IR is growing, although the focus remains firmly on states as the major actors and resolve as the reputational type. For instance, Dafoe and Caughey (2016) considers how US presidents raised in the South have a heightened sense of honor, which interplays with the weight they place on fostering reputations for resolve in interstate conflicts. Other work, such as that of Guisinger and Smith (2002) looks at diplomacy and argues that credibility about resolve arises from honesty, where leaders who keep their word in crisis are trusted more than those who do not. The extensive work of J. Kertzer (2016) shows that reputations for resolve and honor in the domestic realm inform resolve in the international arena. This focus makes sense with the origin of IR in realism, where the systemic conditions of anarchy create a concern for reputations for resolve where security and uncertainty trigger power-seeking behavior and deterrent strategies. However, as Keohane and Nye (2001) make clear, the world is not always governed by high-stake security concerns. In this context, states can seek to acquire meaningful reputations for such disparate and distinct qualities as honesty, violence or hostility, and dealing with secessionists (Guisinger and Smith 2002; T. M. Peterson 2013; Sartori 2016; Walter 2009).

This dissertation is centered around the idea that reputational concerns cause agents to act in sub-optimal ways, that 'crisis' seems critical to when

the costs of reputation increase, and that there is no 'one reputation', most of the time what we have are actors trying to appease multiple audiences at once, making reputational trade-offs between two or more audiences watching them. More importantly, under conditions of anarchy, where enforcement mechanisms are absent or simply ineffective, reputation is perhaps the central disciplinary mechanism that audiences can use to force agents to behave, to force them to comply, and to punish them for non-compliance with specific normative standards.

## 3.1 Crisis

Situations of crisis have played a central role in the literature on reputation for two reasons. First, because the focus of observation has been states or leaders, crisis heightens the influence that state authority and key national decision-makers have over foreign policy (Furniss Jr 1950; P. E. Peterson 1994; Yarhi-Milo 2018). Second, whether it's the dominance of conflict and war, or that of resolve itself as the primary reputation worth fighting for, the situation of crisis bargaining necessitates the role of crisis in heightening the role and costs of reputation. Goldstein (2013, p. 51) defines crisis as "a confrontation between states involving a serious threat to vital national interests for both sides, in which there is the expectation of a short time for resolution, and in which there is understood to be a sharply increased risk of war".[5] This definition distinguishes crises from other situations with slightly different characteristics, or lower thresholds of tension, such as disagreements not of vital interest, and where military force is not relevant, and war is not eminent (Bouchard 1991; Hermann 1969; Hermann 1972; Snyder and Diesing 2015). Resolve plays a pivotal role in the international

[5] See also: Snyder and Diesing (2015, p. 6) "a sequence of interactions between the governments of two or more sovereign states in severe conflict, short of war, but involving the perception of a dangerously high probability of war."

crisis bargaining literature, and "much of the literature... is concerned with conflict situations more generally" (Morgan 1990, p. 312). For conflict bargaining, crisis matters to the study of reputation by bringing into focus and heightening the costs of an actor's decisions. For other types of reputations, however, there is nothing inherent in the logic of reputation that would explain why crisis is necessary to when it does or doesn't matter for how actor's make decisions. Nevertheless, where the focus has been on resolve and crisis, it is not clear whether other types of crisis in international relations affect the centrality of reputation, or the mechanisms through which reputation works. Furthermore, a definition of crisis would necessarily need to be adjusted since the definitions above are explicit about the high probability of war. This dissertation seeks to contribute to this gap by examining how different crisis - geopolitical crisis, economic crisis, and natural disasters, may affect the role of reputation among different groups of non-state actors. There are specific definitions of crisis that fit these three contexts in each paper. But since this dissertation suggests that during times of crisis - such as negative GDP growth, geopolitical conflict, and extreme weather events that affect large demographics - these events *create* reputational crises, it's worth detailing a definition of this. Reputational crisis is "a major event that has the potential to threaten collective perceptions and estimations held by all relevant stakeholders of an organization and its relevant attributes" (Sohn and Lariscy 2014, p. 24).

The three papers in my dissertation explore reputation either in new crisis contexts, or by considering different types of actors and decision-making in these crisis situations. For instance, while my second paper considers a typical conflict situation - Russia's invasion of Ukraine - rather than decisions to retaliate or mobilize, by focusing on firms, the action during the crisis is not

about escalating conflict or resolve, but rather whether or not the firm issues a statement to withdraw. Lastly, work on natural disasters in political science mostly considers the role of this type of crisis in affecting voting behavior (Healy and Malhotra 2009; Gasper and Reeves 2011), conflict and human security (T. Nelson 2010; Mitchell and Pizzi 2021), migration (Mitchell and Pizzi 2021), and foreign aid (Carnegie and Dolan 2021). There has been less attention to considering natural disasters as crisis points, which increase focal attention on climate change behavior and non-state actors decisions. In my third paper, I contribute to this deficit by considering how reputation matters following natural disasters by focusing on how firms make decisions.

# 4 Situational vs Dispositional Reputations

The debate over whether the features of a crisis overwhelm how observers consider historical types of reputations heralds a related and parallel debate over dispositional versus situational traits of reputation.

The dichotomous narrative of reputation is presented to readers as a story of an actor's disposition, or the situational traits that contextualize the behavior and decisions of the actor – meaning that the actor's reputation does not matter. Of course, the inspiration for this duality found its purchase in the work of resolve where the element of crisis in the context of a high-risk issues, such as war, amplifies the situational characteristics and emphasizes past characteristics of an actor's behavior in similar situations. Dispositional characteristics refer to a judgments of someone's character that is then used to predict or explain future behavior.

A dispositional reputation is one where the nature or characteristics of the

individual are reflexively used to explain behavior, or determine how an actor may behave (Mercer 2010). In contrast, a situational attribute would mean that an actor makes decisions based on the transient features of a particular context. This would suggest that reputation does not matter, as I touched on in the literature review. In the context of my three papers, if someone were to adopt a situational perspective, they would suggest that given the circumstances, all actors would a) attempt to design policy instruments and utilize covert instruments to appeal to different groups to secure re-election, b) all firms would withdraw, and c) all companies would avoid disclosing information on their behavior on climate change to avoid scrutiny.

Of course, the variation present in each of these papers would suggest different attributes and features of actors matters here. The other issue is that these actors are behaving in a way that is seeking to preserve, maintain, or defend a reputational type in that situation; the agent has concerns over what reputation will be projected to their audiences after the situation resolves itself or enough time passes; the time horizon is always longer than just the immediate situation.

Other scholars have pointed out the difficulty of disentangling these two items; most notably J. Kertzer (2016) who highlights that dispositional features could be of dispositional origin. Fearon (1995) makes a similar suggestion in the context of war. I will extend his thoughts to the political economy context here; how an actor behaves in a given crisis or domain may be affected by private information, different estimates over how the outcomes of a situation will play out, commitment problems, or indivisible issues. All of these three variables are weighed down by an actor's perception of them, and the meaning the actor places on them. So too, can situational factors change the cost-benefit analy-

sis an actor makes over the trade-offs of acting in their best interests, versus the reputational damage they could incur from doing so. The context of Russia's war on Ukraine and natural disasters induced by climate change shape the preferences of audiences, which in turn places pressure on the company to take measures to either conceal their behavior, or force their hand.

Situation and disposition are likely intertwined, but despite how broad both of these concepts are in traversing how observers make judgments over reputation, the assumption underpinning both of these stories is the same; reputation exists in a vacuum. For disposition, the vacuum is the singular entity of the actor. For situational explanations, the vacuum is the situation in isolation. Both of these approaches sidestep the role of audiences and how reference points help inform how audiences perceive behavior, and in turn, who an actor looks to when making judgments about how their actions or speech will impact their reputation. By privileging the role of audiences, without whom, reputation would be meaningless, I argue that the third dimension of the story of reputation is that it is a *group-based* concept, where reputation is a positional good.

The empirical record on the impact of how actor's use the past behavior of another actor to inform expectations or predictions in contemporary decision-making has been mixed. Proponents include Richard (1981) and Orme (1992) who find that the Soviet Union inferred US levels of resolve based on past inaction. On the other hand, critics including Hopf (1994) have argued that Soviet leaders did not glean insights on US behavior in Europe from past actions in peripheral third world countries. Building on this, Mercer (2010) argued that reputations for resolve were more steadily built by situational characteristics of a given crisis rather than the historical dispositional features of an opposing state. In arguing this, however, Mercer stresses the importance of the relation-

ship between the observer and the state (whether they are allies or adversaries). Thus, while their past behavior in general may not matter, their past behavior in relation to the observer state does. Press (2005) also finds little support for past behavior on reputation, referencing US and German policy-makers during World War II, and argues instead for the primacy of current capabilities and interests in perceptions of resolve. Mercer and Press both privilege realist variables of current power and interest over past behavior in explaining outcomes. However, as Weisiger and Yarhi-Milo (2015, p. 492) argues in finding empirical support for the effects of past behavior on reputation, much of the empirical evidence in past studies can be explained by "focus on crisis decision making, where information gleaned from past action will already have been incorporated into broader estimates of interests and hence is less likely to be referenced directly, as well as from their failure to recognize that reputation acts through estimates of an opponent's interests."

In light of this, some critics have attempted to bridge past behavior with situational factors in explaining the influence of reputation. Cuesta et al. (2015) show that, while individuals use historical actions of another agent to establish reputation, they place an equal weighting on the most recent action of that agent. Reputation, in their paper, is measured by "the fraction of cooperative actions and the last action performed". The underlining issue with this position is how to separate most recent behavior from past behavior; and the responsibility of current administrations for the behavior of prior administrations. This position also assumes complete information on behalf of observers, or that all observers would place equal weight on past actions.

The second debate that underscores this first part of the dissertation is

whether leaders or states develop reputations. When IR scholars discuss reputation they tend to discuss individual leaders' reputations as temporary or distinct from a state's reputation. For instance, Dafoe and Caughey (2016) discuss leaders' concern for honor and how it can lead to a preference for fostering reputations for resolve; Guisinger and Smith (2002) emphasize individual leader honesty and its importance in signaling credibility in diplomacy, and J. Kertzer (2016) asks why some leaders display resolve during confrontations and crises, while others choose not to. Parallel to this is work by Guisinger and Smith (2002), Chiozza and Choi (2003), and Wolford (2007) who emphasize the relationship between tenure and institutions that mold those reputations. In this body of work, reputations of states are divorced from reputations of leaders. Similarly, the current literature on reputation often assumes state preferences to be a homogeneous block – leaving out the burgeoning growth of non-state actors and how they may factor into a domestic governments' concern over reputation, or whether the reputational concerns of firms trumps that of the state in certain domains or international events. In this way, the relationship between the reputation of the individual or group of elites, and the reputation of the collective state itself, is unclear. It's an idea entertained in passing by Jervis (2017) who explores how the reputation of the statesman is used as a focal point for the public to form perceptions of the reputation of the nation state, but ultimately Jervis leaves this idea up to the task of future researchers.

Although many studies have found evidence that leaders can acquire reputations for various qualities including resolve, honesty, violence or hostility, keeping alliance commitments, repaying debts, or even handling secessionist groups (Guisinger and Smith 2002; Gibler 2008; Tomz 2007; Walter 2009; G. D. Miller 2004; T. M. Peterson 2013), this is weighed against work arguing that

reputations are situational and these reputations only matter in particular contexts, and the precarious 'stickiness' of reputation for leaders for certain types of reputations. For instance, Lupton (2020) argued that reputations stick to states primarily in the absence of information regarding how a new leader will behave. Lupton traces how leaders build reputations for resolve during the early period of their tenure in statements or key behaviors - which ground observers expectations of future behaviors. Reputations for cooperation are less likely to stick to leaders when the informational available is much more diluted. This work opens up questions on whether leaders might develop different reputations among constituents based on their mode of entry, and the expectations they set among their audience when they're elected into office.

# 5 Psychology and the role of Group behavior in Reputation

Reputation relates strongly to the notion of identity and group behavior. Group members will impose sanctions on perceived deviants of a particular norm or socially accepted behavior, and "would-be deviants anticipate the sanctions in advance" (C. R. Sunstein 2001, p. 17). The actions of all members of a group provides information about each agent, and each individual's expressive actions generates a reputational externality within the group. Every individual cares about their reputation within the group, and they are incentivized to behave in a way they believe other group members expect them to. Concern for reputation exerts pervasive influence on group behavior (C. R. Sunstein 2008; Anderson and Holt 1997; A. V. Banerjee 1992; Bikhchandani, Hirshleifer, and Welch 1998; Daughety and Reinganum 1999).

Experiments in behavioral economics show that self-interest theory is only predictive of agent behavior in laboratory settings; but in the context of a well-functioning marketplace, reputation alters behavior to align with the predictions of social preference models (List 2006). Reputation has also been shown to resolve the lemon problem; when transactions are repeated, and information to verify quality exists, high quality is provided (Thierer et al. 2015).

The role of reputation in altering group behavior and promoting pro-social outcomes has been found in biological research. Studies of the reef cleaner fish, *labroides dimidiatus*, suggest that clients of the fish observe cleaning sessions and spend more time with cleaner fish thought to be cooperative. When the cleaners feel they are being observed, they behave more cooperatively (Bshary and Grutter 2002). The incentive to maintain cooperative behavior increases with observation (Tennie, U. Frith, and C. D. Frith 2010), and declines with anonymity (Milinski, Semmann, and Krambeck 2002).[6] Observation of an audience increases an individual's attention to an ideal behavioral standard (Carver and Scheier 1981). Developmental psychological work shows that not only do individuals make selective choices on self-presentation to promote their own reputation, they change which characteristics they choose to promote about themselves depending on the preferences of their audience (R. Banerjee 2002). Thus, what is the 'ideal' behavioral standard that promotes the best reputational type changes depending on which category of reputation the audience values the most.

From childhood, people are aware that their own reputation is the product

[6]Milinski, Semmann, and Krambeck (2002) Reputation helps solve the 'tragedy of the commons' and demonstrates the importance of reputation for the emergence of cooperation.

of direct-experience and also misinformation, gossip, and indirect-experience (C. D. Frith and U. Frith 2012). Additionally, agents can place significant weight on second-hand information regarding reputation, with one study showing that even with direct observability of an individual that a participant would rate positively, when exposed to negative indirect information about that same individual, the participant would change their overall assessment of the individual (Sommerfeld et al. 2007). The effects of indirect information on altering direct assessments has been replicated in other studies (Zarolia, Weisbuch, and McRae 2017; Thorsteinson and Balzer 1999). Second-hand information can manipulate reputational types, even if such information is not directly observed. Indirect information can reduce reliance on the feedback mechanism used in repeated games to assess trustworthy reputational types of partners (J. Li, Delgado, and Phelps 2011). The development of a poor reputational type will affect interpretations of future actions, and agents categorized as having poor reputations are less likely to be forgiven (Darby and Schlenker 1989).

As individuals and organizations develop reputational types over-time, observers, clients, and stakeholders use this information to assess which categories the individual or organization falls into. The management and micro-economic literature shows that such reputational categorization affects behavior and possibilities of cooperation. Individuals may choose to only engage in generalized social exchanges if their chosen partners are of a particular reputational type, as Ostrom and Walker (2003, p. 43) explain, "trustworthy individuals who trust others with a reputation for being trustworthy (and try to avoid those who have a reputation for being untrustworthy) can engage in mutually productive social exchanges. . . . So long as they can limit their interactions primarily to those others with a reputation for keeping promises".

Knowledge of reputation can be bifurcated into reputation accuracy and reputation insight. The former refers to the accuracy of which an observer can assess how others view a target's reputation. In contrast, reputation insight is the extent to which the observer is aware of the discrepancy between their perception of the target's reputation and how others perceive the target (Solomon and Vazire 2016).[7] Knowledge of reputation is especially important when making decisions over whether to cooperate; how a target is perceived by the wider group can affect whether an agent's association with the target has a positive or negative affect on their own reputation (ibid.). Knowledge of others' perceptions of reputation influences behavior, where the agent internalizes the prescriptions associated with the behavior in order to reap the rewards.

Internalization has been linked to behavioral prescriptions of identity in psychology, and to following norms in International Relations (Kelman 1958). So far reputational prescriptions have not been linked to a process of internalization. However, the individual gains of conforming to a group reputation suggest the incentives around internalizing prescriptions for identity are similar to those of reputation. Individuals selectively interact with potential partner's on the basis of their association with a group or organization. In some situations a partner's history or character profile is inaccessible, their membership to the group becomes a proxy for their reputational type. For instance, in one game, Gurerk, Irlenbusch, and Rockenbach (2006) show that participants migrate to institutions with sanctioning mechanisms over institutions that lacked punitive measures. It follows that individuals internalize the prescriptions associated with the reputation of their group to boost the prospects of partnerships (En-

[7]See also the literature on third-party metaperception, which explores how one observer thinks another agent perceives a target. This is distinct from reputation insight which refers to many perceptions (Kenny 2004).

gelmann, Herrmann, and Tomasello 2018). One study on e-commerce argues that social networks act as a proxy for reputation; the average reputation of the surrounding nodes in a network is indicative of the reputation of the central node (Hogg and Adamic 2004). Thus, individuals care about the reputation of their group, internalizing the prescriptions that maintain this reputation.

The process of internalization relies on 'identification' whereby the agent learns what behaviors fit the mold of a reputational type, and the behaviors that are incompatible with this type. If an agent within their own reputational mold violates these prescriptions, they created anxiety for other agents in this group by threatening the reputational category for those members. In this way, group behavior has a huge impact on how individuals perceive of their own reputation; or the reputation of the group they either formally or informally belong to. Group behavior is integral to the study of reputation in IR, and below I consider how this logic helps inform the mechanisms of reputation that are discussed in this dissertation.

# 6 Mechanisms of Reputation: Cascades, Externalities, and Early-Movers

The mechanism through which reputation operates has been narrowed significantly in IR through the focus on resolve as the major reputational type. For these scholars, reputation is a mechanism for conveying information regarding a state's propensity for resolve, whether this be defense of honour as in Thucydides classic narrative of the Peloponnesian War, or willingness to retaliate and punish (T. C. Schelling 1956; T. Schelling 1960). Reputation is outlined as a mechanism in itself to encourage deterrence or signal resolve (Crescenzi 2018).

## 6.1 Cascades

Reputations are not formed in vacuums, and they rely on a tenuous informational environment. When a state's reputation is formed, augmented, destroyed or reconfigured in world politics, it is through a deliberative process among groups of individuals and states. I will examine how these informational and deliberative influences can produce 'reputational cascades'. Building on the definition offered by C. R. Sunstein (2019)'s work on social cascades, I define reputational cascades as large-scale judgments and statecraft movements in which actors "end up thinking something, or doing something, because of the beliefs or actions of a few early movers" (C. R. Sunstein 2002, p. 19). The concept of 'early movers' is not mainstream in IR discussions, despite how important it is for shaping policy. Take for example COVID-19; early-movers like China, South-Korea and Japan responded to the COVID-19 crisis with what would become known as 'lockdowns'. China's Hubei province locked-down in January 2020; an action that was swiftly followed by Italy, and a few days later Spain and France. The US's initial warning on travel to China on January 27th 2020 triggered Australia, Singapore, and Russia to ban travelers from China. The initial nature of responses to an unprecedented global pandemic was to watch how the unknown virus played out in first-hit countries, to observe how state's managed transnational and then national movements, and whether and how to implement restrictions. These cascades explain how legal, political and social movements can be rapid and produce tenuous situations.

One of the first studies finding evidence of social cascades was conducted by Salganik, Dodds, and Watts (2006). In their study, participants would hear and download one or more of 72 songs and individuals were not told what anybody else had downloaded or their preferences. Eight other subgroups were created

where individuals could view how many times each song was downloaded. The findings showed that the songs ranked best or worst by the control group never ended up at the top or bottom of the other subgroups, instead if a song received an early burst of downloads, it excelled in the rankings. Without the early movers pushing the song, any other song could fail (Salganik and Watts 2008; Salganik and Watts 2009). The effect of cascades extends to interpretations and judgments over facts as well. The judicial system of basing judgments on precedents functions as an informal institution of cascade effects, where decisions are based on the verdict of others rather than an individual judgments (Daughety and Reinganum 1999). An important study by Miceli and Coşgel (1994) shows how a judge's concerns for reputation informs, restrains and inspires judgments. In the authors' model, the preferences of judges were represented by a utility function with two components: private utility and reputational utility. The judge's private utility arose from his personal view of how a case should be decided, while his reputational utility depended on his expectation of how his decisions would be viewed by observers of the legal process: other judges, lawyers, litigants and the public.

A reputational cascade is not a normative concept in and of itself, however they can lead to "widespead errors, factual or otherwise. Cascades need not involve deliberation; but related problems infect processes of group deliberation, as we will now see" (C. R. Sunstein 1999, p. 9). Rather convergence based on early-movers can lead actors to form distorted perceptions of another actor's reputational type, leading to inefficient outcomes (at best) and escalatory high-risk situations (at worst). In some situations of IR a cascade effect will overcome collective torpor or reluctance in creating public goods, for instance acting on climate change. In other situations cascade effects will produce large-

scale distortions in state-elite judgments, creating unjust and escalatory foreign policy responses as state's quickly align their public beliefs and positions with 'first-movers'. Another reason why the phenomenon of 'Reputational Cascades' occurs in IR is the high prevalence of risk (Kuran and C. R. Sunstein 1998).

Reputational cascades are a distinct phenomenon to informational cascades, but we can think of them as siblings (Kuran 1998).[8] Here is the distinction: in an informational cascade, individuals fail to reveal private information to the group out of concern that their information is 'wrong' if their information contradicts that of the first early-movers. In contrast, in reputational cascades the failure to reveal private information to the group does not arise out of a fear of that information being 'wrong' but rather the fear of exclusion or disapproval for contradicting the group. In order to preserve or project a particular reputation, individuals will speak out - or remain silent (C. R. Sunstein 1999, p. 8). Kuran (1998) refers to this as 'pluralistic ignorance'.[9] An example of pluralistic ignorance and its relationship to reputation in creating a cascade effect is exemplified by US attitudes to segregationist policies in the 1960s and 1970s (O'GORMAN 1975; O'Gorman and Garry 1976). A series of studies showed that Americans largely overestimated white support for forced racial segregation, while just 18%

[8] Although conceptually similar, reputational cascades are distinctive from that of the 'domino effect'. The domino effect refers to, "changes in one country's political institutions spread to neighboring countries, affecting these countries' political institutions similarly, which spreads to their neighbors, and so on. According to the democratic domino theory, for instance, increases in one nation's democracy lead to increases in its neighbors' democracy, leading to increases in their neighbors' democracy, and so on" (Leeson and Dean 2009, p. 534). Reputational cascades are distinctive in several ways. The domino effect refers to observed outcomes but has no mechanism - instead there are several speculated mechanisms that explain why, for example, neighboring countries would 'fall' to communism in succession (Simmons, Dobbin, and Garrett 2006; Pevehouse 2002a; Pevehouse 2002b). In contrast a reputational cascade is a mechanism by definition. Second, the cascade effect is driven by concern for reputation, while the literature on the domino effect has never included reputation as a possible mechanism. Third, the logic of both are different: in a domino effect, the institutional structure of the actor changes not because of concern for observers but because of internal preferences, in a reputational cascade, actors are making decisions based on a) peer behavior and b) concern for how observer's view their reputation.

[9] For more on pluralistic ignorance see: J. Shamir and M. Shamir (1997), D. T. Miller and Prentice (1994), and J. O'Gorman (1986).

privately supporting forced segregation, 47% of respondents believed that the majority of whites supported forced segregation. This perception of majority opinion created a willingness to endorse racial segregation, even if it conflicted with the individuals personal preference (Kuran 1998). Experimental studies have largely confirmed pluralistic ignorance in many settings, D. T. Miller and L. D. Nelson (2002) finding pluralistic ignorance motivated voters who voted for Al-Gore, not because he was their preferred candidate, but because of their aversion to Bush, to believe that all other voters strongly preferred Bush. Other studies have uncovered pluralistic ignorance in alcohol consumption (Schanck 1932), liberal beliefs of university students (Korte 1972), the media and journalism (Gunther and Chia 2001), and even a willingness to wear a sandwich board (Ross, Greene, and House 1977). Of course in all of these instances we can not a-priori rule out that an individual might be motivated by genuine preferences rather than reputational pressure.

Another example of a reputational cascade might be communism in Eastern Europe, where it was at least partially sustained by the incorrect belief by individuals that the majority of people supported the existing regimes (Kuran 1998). In extreme cases of pluralistic ignorance, unpopular private views become 'unthought'. The term 'unthought' was used by the Islamic scholar Muhammod Akroun to describe thoughts one can not admit having for fear of severe reputational damage, and Kuran (ibid.) builds on this idea to argue that public pressure can lead to knowledge falsification through the transition of 'unthinkable' thoughts to 'unthought' thoughts. For civilians, the idea that 'freedom of speech' insulates them from this process may not be well-founded, even in the absence of legal ramifications, unpopular opinions generate a social penalty, damage reputations, and lead to other sources of loss (jobs or social

groups). The secret ballot, in contrast, is an effective way to overcome pluralistic ignorance.

A famous laboratory experiment on urns illustrates the effects of reputational cascades (Anderson and Holt 1997). In this experiment, subjects had to guess whether the experiment was using an urn filled with two red balls and one white ball (urn A) or one containing two white balls and one red (urn B). For each period one of the urn's balls were emptied into a container, a randomly selected participant selected one ball, recording the ball's colour and their opinion about which urn was in use. The subject's colour was undisclosed to the group, but the subject's opinion on which urn was in use was shared. All participants would complete the same task and after the last participant, the answer of which urn was in use would be announced. The experiment revealed that individuals would make decisions about which urn was in use based on the majority of previous announcements, even if it contradicted with the color of the ball (Hung and Plott 2001). Over 77 percent of each round of the experiment revealed cascades, with few 'private signals', that is information provided by the individual's own draw. Interestingly, when one person's ball draw contradicted the decision of two or more previous participants, the next participant was likely to ignore this 'outlier' and follow the consensus of those who went before.[10] Although the majority of decisions followed Bayes law, the conclusion from the study was, "Initial misrepresentative signals start a chain of incorrect decisions that is not broken by more representative signals received later" (Anderson and Holt 1997, p. 847). Hung and Plott (2001) replicated a version of this experiment using rewards where subjects were paid \$2 for a correct group decision, and fined \$2 for an incorrect group decision, with the group decision based on the majority

[10] 72 percent of subjects followed Bayes's rule in the Anderson and Holt (1997) experiment, and 64 percent in Willinger and Ziegelmeyer (1998).

and no individual reward for individual decisions. In this experiment, just 39 percent of rounds witnessed cascades, and in 92 percent of cases, individual announcements matched their private ball draw. The group reward encouraged individuals to reveal private information. In the first experiment, participants did not care about sending an accurate signal to others, it was essentially an "informational externality" which did not affect an individual's gain, if the signal misled others, the participant had no real reason to care.

Another example of a reputational cascade is given by C. Sunstein (2019, p. 68):

> "Suppose that Albert suggests that genetically modified foods are a serious problem and that Barbara concurs publicly with Albert, not because she actually agrees with Albert but because she does not wish to seem, to Albert, to be ignorant or indifferent to human health and environmental protection. If Albert and Barbara agree that genetically modified foods are a serious problem, Cynthia might not contradict them publicly and might even seem to share their judgment, not because she believes the judgment to be correct but because she does not want to face the hostility or lose the good opinion of others. It is easy to see how this process might generate a cascade. Once Albert, Barbara, and Cynthia offer a united front on the issue, their friend David might be most reluctant to contradict them even if he thinks they are wrong."

We can see how this would play out in IR. Suppose a state suggests that the actions of state B in dumping toxic nuclear waste is a serious problem, and that

state C concurs publicly with state $A$, not because state $C$ necessarily agrees with state $A$ but because they do not want to appear to state $A$ as ignorant or indifferent to health or environmental protection. This effect is even greater for international organizations - for instance, thinking through the above scenario but replacing state $A$ with an IO. The cascade becomes clear if state $D$ and $E$ become more reluctant to contradict the united front of state $A$ and $B$, even if state $D$ or $E$ believe privately (and individually) that they are wrong.

Of course in world politics with deliberative decision-making by states and actors, other states are uncertain whether public statements are the product of independent knowledge, informational cascades or pressure induced by reputational concerns.[11] Legislators suffer from reputational cascades when they're pressured by party-members or the public, for instance consider the Conservative parties decision to hold a referendum of Brexit. Although we know there were few rebels who defected from the party, many more publicly supported Brexit despite private misgivings in order to avoid injuring their reputations.[12]

An early paper by Akerlof (1976) argues that the real world does not resemble the Arrow-Debreu utopia (a general equilibrium model of perfect competition) because information is both costly and incomplete, leading to the use of indicators to make predictions of individual or economic behavior. There are different types of indicators, but indicators with social origin can lead an economy into a low-level equilibrium trap. Reputational judgments distort indicators by leading outsiders to assess insiders based on the judgments others have made of the organization, institution, or group they belong to. Signals in IR function as

[11] For corporate scandals, see: Thompson (2004).

[12] David Cameron himself admits he regrets how he handled the Brexit campaign, influenced by his concerns for his reputation within the party (with a vocal right-wing faction) and the public (with growing anti-European sentiments). See Also: Stimmer and Gliserman (2022).

indicators, although the IR literature does not currently differentiate between types of signals, most signals sent have heavy social dimensions. Furthermore, information in world politics is both incomplete, uncertain and costly if assumed true.

Reputational cascades as a mechanism of reputation have largely been absent from the literature in IR, despite the profound applicability. Exceptions include very recent work by Hale (2022) to explain the rally around the flag affect around Putin, and Pearlman (2018) protests in Syria. Considering reputation as group-based, it is surprising that so few works have considered the mechanism through which reputational pressure creates within-peers effects, such that there is a contagion effect of reputation. Outside of political science, and the work of Sunstein himself, almost no work has considered companies as having network effects of reputation and the effect of reputational cascades on behavior. My second (2) and third papers on company behavior in Russia, and disclosures on climate (3) make direct contributions to the study of reputation in IR, by introducing this novel concept and suggesting that these cascades explain how reputational concerns are both induced by, and alter, group-dynamics, leading to large-scale change. Below, I consider how this mechanism is triggered, through the role of early-movers.

## 6.2 Externalities and Network-Effects

In this dissertation, I introduce the notion of reputational externalities to explain how an agent's behavior, who subscribes to a particular reputational category, falls outside the prescriptions set by that reputational category can threaten the collective reputation of the group . We can conceive of reputational externalities to be a situation where "the actions of one group may damage the reputation of

another group, thereby reducing overall welfare" (Evans and Guinnane 2006). A historical economics example of this effect is illustrated by Evans and Guinnane (ibid.) in 19th century German credit cooperatives. These cooperatives were divided into an urban group, led by Hermann Schulze-Delitzsch, and a rural group, led by Friedrich-Wilhelm Raiffeisen. Although the cooperatives served the same function and shared many similar features they differed in their size, liabilities, lending practices, and financial relationships with other institutions. However, because the credit cooperatives fell under the same umbrella among the German public, when news emerged of bad practices or poor outcomes from the Raiffeisen cooperatives, the Schulze-Delitzsch cooperative's reputation was negatively impacted. This led to the latter group lobbying for regular audits of all cooperatives. This phenomena is discernible in IR; for instance, when states join together under an organization or grouping, the behavior of one agent can damage the reputation of the other agent. Or, perhaps, when heterogeneous agents or bodies develop common reputations through associations or long-standing alliances. On a more individual level, we know reputational externalities occur for leaders within political parties and leaders can be ousted if they threaten the reputation of the party. This type of reputational externality is tested in my third paper on firms who may disclose to avoid the negative externality of belonging to a group that has developed a poor reputation on climate change.

# 7 Reputation as a Dynamic Group Effect

I make three broad contributions to the literature on reputation in International Relations in this dissertation. The first is that audiences are diverse and operate within a group-dynamic; they will look to each other when forming judgments

about the signals sent by actors through behavior and decisions. They will also take cues from each other to guide the way in which they voice their conclusions over that actor's reputational type. The second is that there are multiple audiences at any given time that may place greater or lesser weight on certain indices of reputation that may conflict with each other. The third is that reputations are interdependent; the actions of other members of a group take in a given situation can affect a (formal or informal) member of that group, this will in turn determine how observers cast judgments over the reputational type of the actor. By pushing back on the necessity of history-dependence for judgments of reputation to be made, I instead appeal to a group-based conceptualization of reputation that considers how observers make judgments based on peer behavior. First, however, I consider the literature on reputation as group-based within the hemisphere of IR, before I show that the broader field of reputation outside of this prism largely considers it a group-based concept, and make the case that IR would benefit from considering reputation in this way.

There are few scholars who consider reputation a group-based good in IR. One such work that comes close to considering reputation as group-based is Crescenzi (2018) work on state reputations and alliance formation. Crescenzi (ibid.) argues that states make decisions on whether to go to war, or form an alliance, based on the reputation of other states with each other.[13] His argument is that states form inferences based on the interactions of states with other states within a system, and through this states learn information that they use to form judgments over an actors reputational type (for Crescenzi, this is credibility in keeping commitments, and resolve for peace or alliance formation). Temporal proximity is also important; the closer the event that was used to

[13]See also: (Henry 2022) on alliance interdependence, although Henry differs here by convincingly arguing that this results from developing a reputation for 'observed reliability', not resolve or royalty.

generate information on the reputational type of the agent, the greater weight applied by the observing state. Crescenzi uses an illustration of state interactions between North Korea, the United States, and Iraq. Accordingly, North Korea uses the history of US interactions with Iraq as information on reputation.

The greater the similarities between states in terms of policy and power resources, the more the state is able to use another state as a proxy for its likely relationship with states in a system. The inter-relationships of states used to inform reputation is interesting in this context, although Crescenzi (2018) does not go further in his argument that state's use alike proxies to make informed decisions. What he is missing here is that state's use benchmarks on behavior to help weigh indices which in turn are used to form judgments on behavior; those benchmarks are decided by the within-group behavior. Groups are not necessarily fixed, and their formation might depend on the context or the situation, however they are formed by alike actors who have something in common. These actors may be facing a similar problem of which they must all make a decision; or an actor has positioned themselves within this group by joining it or establishing themselves in opposition to this group. In both scenarios, the pre-existing status quo of the group has now set the benchmark of appropriate behavior that the audience will accept. When one actor within the group departs from the status quo, it resets the benchmark, forcing the actor's within the group to make choices which will inform audiences of their reputational type. In my second and third paper, the group is formed by firms who face the same decision. For paper two this is to leave Russia, for paper three this is to disclose on climate change. Audiences use heuristic short-cuts to assign these groups; industry and sector-level characteristics are the most useful ways of comparing firms for consumers.

When early-movers within these sectors depart from the status quo and leave or issue disclosures, they reset the benchmark of acceptable behavior among the audience. Audiences now form different expectations over the range of possible actions open to the actors in the group, and therefore actors that do not adjust to the new status quo sends a signal to the audience that these actors are a different type. In doing so, they set the benchmark of appropriate behavior.

Work on status has approached the role of interdependence of reputation, where scholars have considered how group membership conditions status. For instance, Duque (2018) borrows from Weber (1978) to argue that states achieve status through gaining admission to a club through following the rules of the in-group; therefore a state's relationships with other states influence its ability to achieve status. This work is focused on status, and it concerns formalized and institutionalized rules of behavior, rather than the norms of behavior, where the institution sets and governs these rules rather than the states themselves. From this vantage point, reputation is a positional good where the rewards or benefits of maintaining a particular type of reputation are accrued not from how well the actor has performed in absolute terms, but rather the comparison of how the actor's behavior compares to their peers.

Existing studies on reputation and/or status do not clearly distinguish between status and reputation. In fact, several IR books published on reputation do not engage at all with conceptual differences (Mercer 2010; J. Kertzer 2016; Crescenzi 2018; Tomz 2012), and even some articles on status fail to distinguish it from reputation (Wohlforth et al. 2018). One difference is perhaps found in measurement. Traditional notions of status are almost always about capa-

bilities and how observer's ranking your material status and power in relation to other states. Dafoe, Renshon, and P. Huth (2014) defines it as "attributes of an individual or social roles, especially those attributes related to position in a deference hierarchy", and many quantitative studies on reputation control for status by using the COW on major, middle, and minor powers (Crescenzi 2018). However, socially constructed notions of status, on the other hand, are directly related to group-based notions of reputation. After all, by moving position within a group where your reputational type is a positional good vis-a-vi your group members, you're afforded a better status. Arguments that have been made to differentiate social status and reputation are weak. Take for example this argument:

> "One feature distinguishing them seems to be the degree to which they are in the power of the actor or community. Reputation and prestige are more often regarded as being somewhat in the control of the actor: actors can seize, acquire, and invest in their reputation and prestige. Status, on the other hand, is more often regarded as a function of the community: it is granted or accorded by others" (Dafoe, Renshon, and P. Huth 2014, p. 376).

This argument is difficult to maintain since both status and reputation are influenced by the actor, and are equally observer-dependent - it's also the case that investing in reputation improves status. This point is made by Yarhi-Milo (2018), who presents a more convincing argument for the difference; in her model reputational types help build and improve status:

> "Indeed, reputation for resolve and status are intimately linked in the consequences of international crises: a reputation for greater or

> lesser resolve will influence a state's position in a deference hierarchy. Similarly, status influences reputation by setting expectations of behavior: one's gain in status 'will color how others talk about and perceive one's past actions'" (Yarhi-Milo 2018).

Thus status-seeking is a motive for reputation-building behavior. Another differences can be found in the logic of status, which only tells us that states maneuver to gain advantage over each other, rather than to avoid social penalties. In the status literature, hierarchy is the point. In contrast, for reputation, there's the possibility that all members of the collective have the same reputation, because all members have taken similar actions. Indeed, members within the collective itself help establish the group's reputation.

This is a point, interestingly, that is made in the economics literature on distinguishing between reputation and status: "If one organization loses status relative to others, then it follows logically that those other organizations must increase in their relative status. As a result, status is inherently hierarchical. The same is not true of reputation. If all firms had the same flawless record of paying their suppliers within 30 days, they would all have equivalent reputations for being good credit risks" (ROLFE 2012). Thus, while reputation can be a positional good, there is also a scarcity element of status that is not present in reputation. An additional difference highlighted in the literature include types - reputations can develop for specific traits, while status arises "from factors that are more numerous and diffuse" (Bothner, Godart, and Lee 2010, p. 8). Although the economics literature seems less concerned about distinguishing between the two, with some scholars going so far as using them as synonyms.[14]

[14]See Burt (2005, p. 134) who uses the term reputation to describe Podolny (2010) work, even though Podolny explicitly distinguishes between status and reputation.

# 8 Dueling Audiences

Defining reputation as a dynamic concept that fluctuates in reaction to peer-behavior has two implications; reputational externalities for collective reputations, and reputation stealing externalities for individuals. What this comes down to, then, is that reputation is not constant, it is not a single entity that can be viewed the same from all vantage points. Reputations are structures that are constantly changing, depending on the vantage point of the viewer. Actors develop different reputational types depending on the vantage point of the observing actor (Tversky and Kahneman 1974), such that state B and C could simultaneously believe A to be both a 'good' type and a 'bad' type if their vantage points on that behavior differed. Based on the four issues with the premise of reputation highlighted here, my dissertation would suggest re-conceptualizing reputation to be 'a consensus in judgments formed by a group of an actor's disposition and character based on peer-comparisons, used to explain current behavior and predict future behavior'. The idea of vantage points and network effects brings the focus of attention to an additional contribution of this dissertation; dismantling the idea of a single, homogeneous audience, and introducing the notion of multiple audiences who made hold different expectations over an actor's reputation.

Past work on reputation has assumed an amorphous single audience (or group) places costs on an actor for failing to either live up to their obligations, show resolve, comply with international obligations, or just violate their reputational type. Audience costs are measured through a penalty imposed on the actor, or the actor is assumed to behave in a way that reduces the costs they anticipate they will incur from damaging their reputation among a single audience. One such work that considers the role of multiple audiences is Clare

and Danilovic (2010). These authors take aim at T. Schelling (1960)'s 'indiscriminate reputation' building and multiple audiences. Their study suggests that the presence of multiple rivals (as audiences) incentivizes states to signal reputation. Their study does not consider the idea that audiences may have different preferences over reputational types, that audiences may be located at different vantage points across time and space, such that it affects how they perceive indices and signals of behavior, and how actor's manage reputations with dueling audience expectations. As is the case when the focus is resolve and the audience is international, the audiences in their model are antagonistic and adversaries.

# 9 Firms and Leaders: Non-State Actors with Reputations

The overwhelming focus in the IR literature on reputation is zeroed in on states as the primary unit of analysis (Goldsmith and Posner 2005; Simmons 2000; Hathaway 2002; Hathaway 2005; Tomz 2007). In this sense, non-state actors are a newer group of global actors that IR has only recently begun to consider as major players; this is especially true in the study of reputation. Current studies include Jensen and Malesky (2018) who examine non-state actors and compliance with international agreements through an analysis of the OECD anti-bribery convention. The authors are still focused on the role of state reputation in informing the behavior of firms; the author's argue that agreements use reputational concerns to raise the expected costs of malfeasance for mutlinational corporations. Other work includes, Thrall (2021) who examines where reputation can induce firms to comply with regulatory bodies. For environmentalism in IR specifically, Wapner (1996) shows that activist campaigns, includ-

ing boycotts organized by NGOs, can induce high reputational costs for firms by threatening the legitimacy of the brand's reputation and leading companies to support regulations to hide from the public backlash. Outside of these key works, the literature on reputation in IR with firms as the major actors is sparse. However, moving outside of the prism of IR, the literature on corporations and reputation is vast. M. S. Baucus and D. A. Baucus (1997) explores how multinational corporations who develop poor reputations as rule breakers in international environmental or financial transactions face acute fiscal consequences.

While the business, micro-economic and organizational psychology literature offers several unique avenues into exploring reputation in the context of firms operating within global politics; there are several short-comings to how reputation is measured in this literature. Below I set out the existing measures of reputation, and suggest where the issues with this literature are, before introducing a new measure of reputation for firms in IR.

The business and finance literature often uses ESG scores as a proxy for firm reputation. This measurement choice is based on the intuition that firms build up social capital over time, and that this measurement represents the history of firm behavior; firms with higher scores have more to lose than firms with lower scores. In some ways ESG is a history-dependent definition of reputation. ESG scores are computed annually, not over decades. They're often a combined score on multiple dimensions, and bad behavior on one category can be canceled out by good behavior on another.[15] These scores are tallied over the course of a

[15]There are disaggregated measures of 'environment' scores for firms within that ESG ranking, however they are rarely used in isolation as a reference to the firm, and the same issue arises where the criteria used to generate this score is often not in line with how regular consumer audiences regard these companies behavior in highly salient events.

year, not in response to crisis. In contrast, the IR literature often treats state resolve during a crisis as a marker of a state's reputation; rather than the state building up a measure of resolve in non-crisis situations over the course of a year. Audiences are primed to care about the current cohort of firm behavior in response to a situation, rather than reflecting on firm's past behavior in different situations. Thus, in the case of a genuine geopolitical or environmental crisis, it makes little sense to use ESG as a marker of firm reputation, rather than evaluating the situational features and pressure exerted on firms by their competitors which may generate reputational costs.

Existing research has found ESG scores as poor predictors of firm decisions to withdraw (Ahmed et al. 2023; Lu, Huang, and S. Li 2022). Past research on ESG as a signal of socially conscious corporate behavior and actual ESG-related output has produced mixed findings. Some studies find a negative relationship between ESG scores and firm compliance over carbon emissions or meaningful sustainability initiatives (Raghunandan and Rajgopal 2022; Drempetic, Klein, and Zwergel 2020; Lopez-de-Silanes, McCahery, and Pudschedl 2020).

Three main weaknesses in existing research preclude a more robust understanding of the role of reputation in inducing firm compliance, timing of action, and how firms that remain strategically avoid or mitigate reputational costs. First, reputational concerns on firm corporate responsibility have largely been focused on environmental issues, given the lower visibility of peripheral conflicts among Western audiences, and previous large-scale conflicts occurred at a time when limited agency was assigned to private actors. The politicization of corporations is a relatively recent phenomenon, with the firms existing exclusively as 'profit maximizers' the dominant discourse until recently. Related to this,

consumers had few platforms to foster collective action against multiple corporations. Informational asymmetries were higher, and the costs of mobilizing in the absence of an online platform were also higher. Second, reputational cascades have been overlooked in favour of informational cascades – the idea that a corporation could be making a costly movement in reaction to competing firm behavior has been framed as firm A leaves and sends a signal to firm B that firm A has information about the benefits of leaving/costs of staying, leading firm B to imitate firm A. This idea that firm B is only moving because firm A supposedly has private information on the financial costs of remaining overlooks a related, but distinct, phenomenon; that firm B has 'moved the needle' of appropriate behavior and firm A must now mimic firm B or risk the consequences.

Lastly, the analysis on reputation often assume that both reputational costs are homogeneous, and that some firms are insensitive to such costs. MNC's often face multiple audiences, each of which they must maintain or protect reputations with; at some points in time, as in the Russian invasion of Ukraine, the firms reputation with these audiences can come into conflict. In this situation, some firms choose to throw their proverbial eggs into one audience, and protect one reputation, while other firms attempt to balance both at once. In my dissertation, I show evidence of this by examining how firms react to withdrawing from Russia or disclosing on climate. In this way, my dissertation shows the strategies employed by firms anticipate backlash and reputational costs from audiences.

# 10 Incremental versus Heroic Sanctions

The last literature this dissertation contributes to is that on how reputational penalties are paid - and the expected costs of failing to live up to a reputation, or developing a bad reputation. When reputations for resolve are spoken about in IR it's almost always hinged on a large event where the reputation is 'make or break'. For instance, in the realm of resolve the big event is war or initiation of conflict. However, the idea that the action itself need not be a singular, large event seems to be absent from the literature. In his essay on Bargaining, T. C. Schelling (1956) suggests that in order for an agent to stake his reputation on threat in a credible manner, the agent must engage in a series of smaller, but related, activities that reinforce that the agent will carry out the threat. Schelling states that, "In order that one be able to pledge his reputation behind a threat, there must be continuity between the present and subsequent issues that will arise. This need for continuity suggests a means of making the original threat more effective; if it can be decomposed into a series of consecutive smaller threats, there is an opportunity to demonstrate on the first few transgressions that the threat will be carried out on the rest. Even the first few become more plausible, since there is a more obvious incentive to fulfill them as a 'lesson'"(ibid., p. 297). Schelling is proposing that smaller actions in the lead up to the event are constitutive of the actor's reputational type. One of the implications of this type of thinking, then, is that sanctions may not be absolute, but rather incremental.

Another example of this is evident in Andrew Guzman's contribution *How International Law Works*, international law influences state behavior through the function of state reputations. State's that foster good reputations are rewarded by the benefits of joint cooperation - lengthening the shadow of the

future, lowering transactional costs, pooling resources, and creating common forums of debate (Guzman 2008). States that behave badly are punished by other states by invoking reputational capital. For Guzman, the reputational cost is paid for a lack of follow-through; states that do not honour their word through repeated interactions are punished.[16] Guzman bridges the gap with rationalist or realist scholars who might suggest states only comply when it is in their interest by considering reputation a primary driver of state interest.

These examples are excellent at showing the effect of reputation on conditioning actors behavior - either guiding their action, or explaining why an actor might be sanctioned - however they also obfuscate instances where actors are guided by smaller and less absolute costs, that accumulate over time. Costs are incurred by political leaders, not just through elections held at various time points, but also through shifts in public support between elections. Similarly, costs are felt by companies, not just through cuts in profit, but through slander or misinformation that may affect their brand image - foretelling some future effect on profit even if it's not realized immediately. What I mean by this is that there are 'incremental sanctioning' effects (Coleman 1994), whereby an actor known to have a particular reputational type, commits action X that violates this type, and certain members of the audience impose smaller retaliatory costs on the actor, however these costs are additive and contagious and ultimately have a larger reputational cost on the actor. This idea of 'incremental sanctioning' in the sociological literature demonstrates how norms are enforced, not through one large sanctioning effort, but by smaller efforts over time (ibid.).

The costs are not absolute as IR has suggested, for instance the leader is

[16] See also Tingley and Walter (2011) Actors pay a reputational cost because they fail to live up to their word. But a reputation for honoring one's word is only one kind of reputation at stake in repeated interactions and may not be the most important for all types of reputations.

voted in or out of office, the state becomes known as an unreliable actor, or a liar, or states stop attempting to cooperate with the actor after one demonstration of non-compliance, defection, or cheating. In reality, the costs are usually incremental. As I argue in these papers, the size of the costs usual results from the network of individuals, or the size of the audience, that collectively react to the actor's violation of their reputational type. Just as the network of peers helps determine the reputation of the actor, so to does the network of the audience determine the costs imposed.

# References


Ahmed, Daniyal et al. (2023). "Are ESG Ratings Informative About Companies' Socially Responsible Behaviors Abroad? Evidence from the Russian Invasion of Ukraine". In: *Accountability in a Sustainable World Quarterly, Forthcoming*.

Akerlof, George (1976). "The economics of caste and of the rat race and other woeful tales". In: *The Quarterly Journal of Economics* 90.4, pp. 599–617.

Anderson, Lisa R and Charles A Holt (1997). "Information cascades in the laboratory". In: *The American economic review*, pp. 847–862.

Banerjee, Abhijit V (1992). "A simple model of herd behavior". In: *The quarterly journal of economics* 107.3, pp. 797–817.

Banerjee, Robin (2002). "Audience effects on self-presentation in childhood". In: *Social Development* 11.4, pp. 487–507.

Baucus, Melissa S and David A Baucus (1997). "Paying the piper: An empirical examination of longer-term financial consequences of illegal corporate behavior". In: *Academy of management journal* 40.1, pp. 129–151.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch (1998). "Learning from the behavior of others: Conformity, fads, and informational cascades". In: *Journal of economic perspectives* 12.3, pp. 151–170.

Bothner, Matthew S, Frederic C Godart, and Wonjae Lee (2010). *What is social status? Comparisons and contrasts with cognate concepts*. Tech. rep. Working Paper, European School of Management and Technology, Berlin.

Bouchard, Joseph F (1991). *Command in Crisis: Four Case Studies*. Columbia University Press.

Bshary, Redouan and Alexandra S Grutter (2002). "Asymmetric cheating opportunities and partner control in a cleaner fish mutualism". In: *Animal Behaviour* 63.3, pp. 547–555.

Burt, Ronald S (2005). *Brokerage and closure: An introduction to social capital*. Oxford university press.

Carnegie, Allison and Lindsay R Dolan (2021). "The effects of rejecting aid on recipients' reputations: Evidence from natural disaster responses". In: *The Review of International Organizations* 16, pp. 495–519.

Carver, Charles S and Michael F Scheier (1981). "The self-attention-induced feedback loop and social facilitation". In: *Journal of Experimental Social Psychology* 17.6, pp. 545–568.

Casler, Don, David Ribar, and Keren Yarhi-Milo (2023). "The Many Faces of Credibility: Hawks, Doves, and Nuclear Disarmament". In: *Security Studies* 32.3, pp. 413–445.

Chiozza, Giacomo and Ajin Choi (2003). "Guess who did what: Political leaders and the management of territorial disputes, 1950-1990". In: *Journal of Conflict Resolution* 47.3, pp. 251–278.

Clare, Joe and Vesna Danilovic (2010). "Multiple audiences and reputation building in international conflicts". In: *Journal of Conflict Resolution* 54.6, pp. 860–882.

Coleman, James S (1994). *Foundations of social theory*. Harvard university press.

Crescenzi, Mark JC (2018). *Of friends and foes: reputation and learning in international politics*. Oxford University Press.

Cuesta, Jose A et al. (2015). "Reputation drives cooperative behaviour and network formation in human groups". In: *Scientific reports* 5.1, p. 7843.

Dafoe, Allan and Devin Caughey (2016). "Honor and war: Southern US presidents and the effects of concern for reputation". In: *World politics* 68.2, pp. 341–381.

Dafoe, Allan, Jonathan Renshon, and Paul Huth (2014). "Reputation and status as motives for war". In: *Annual Review of Political Science* 17, pp. 371–393.

Danilovic, Vesna (2002). *When the stakes are high: Deterrence and conflict among major powers*. University of Michigan Press.

Danilovic, Vesna, Joe Clare, and Colin Tucker (2020). "Reputation in Deterrence and Decision-Making". In: *Oxford Research Encyclopedia of International Studies*.

Darby, Bruce W and Barry R Schlenker (1989). "Children's reactions to transgressions: Effects of the actor's apology, reputation and remorse". In: *British Journal of Social Psychology* 28.4, pp. 353–364.

Daughety, Andrew F and Jennifer F Reinganum (1999). "Stampede to judgement: persuasive influence and herding behavior by courts". In: *American Law and Economics Review* 1.1, pp. 158–189.

Drempetic, Samuel, Christian Klein, and Bernhard Zwergel (2020). "The influence of firm size on the ESG score: Corporate sustainability ratings under review". In: *Journal of business ethics* 167, pp. 333–360.

Duque, Marina G (2018). "Recognizing international status: a relational approach". In: *International Studies Quarterly* 62.3, pp. 577–592.

Engelmann, Jan M, Esther Herrmann, and Michael Tomasello (2018). "Concern for group reputation increases prosociality in young children". In: *Psychological Science* 29.2, pp. 181–190.

Evans, Robert and Timothy W Guinnane (2006). "Reputational Externality and Self-Regulation". In.

Fearon, James D (1995). "Rationalist explanations for war". In: *International organization* 49.3, pp. 379–414.

Frith, Chris D and Uta Frith (2012). "Mechanisms of social cognition". In: *Annual review of psychology* 63.1, pp. 287–313.

Furniss Jr, ES (1950). *Congress and Foreign Policy.*

Gasper, John T and Andrew Reeves (2011). "Make it rain? Retrospection and the attentive electorate in the context of natural disasters". In: *American journal of political science* 55.2, pp. 340–355.

Gibler, Douglas M (2008). "The costs of reneging: Reputation and alliance formation". In: *Journal of Conflict Resolution* 52.3, pp. 426–454.

Goldsmith, Jack L and Eric A Posner (2005). *The limits of international law.* Oxford University Press.

Goldstein, Avery (2013). "First things first: The pressing danger of crisis instability in US-China relations". In: *International Security* 37.4, pp. 49–89.

Guisinger, Alexandra and Alastair Smith (2002). "Honest threats: The interaction of reputation and political institutions in international crises". In: *Journal of Conflict Resolution* 46.2, pp. 175–200.

Gunther, Albert C and Stella Chih-Yun Chia (2001). "Predicting pluralistic ignorance: The hostile media perception and its consequences". In: *Journalism & Mass Communication Quarterly* 78.4, pp. 688–701.

Gurerk, Ozgur, Bernd Irlenbusch, and Bettina Rockenbach (2006). "The competitive advantage of sanctioning institutions". In: *Science* 312.5770, pp. 108–111.

Guzman, Andrew T (2008). *How international law works: a rational choice theory.* Oxford University Press.

Hale, Henry E (2022). "Authoritarian rallying as reputational cascade? Evidence from Putin's popularity surge after Crimea". In: *American Political Science Review* 116.2, pp. 580–594.

Hathaway, Oona A (2002). "The cost of commitment". In: *Stan. L. Rev.* 55, p. 1821.

Hathaway, Oona A (2005). "Between power and principle: An integrated theory of international law". In: *The University of Chicago Law Review*, pp. 469–536.

Healy, Andrew and Neil Malhotra (2009). "Myopic voters and natural disaster policy". In: *American Political Science Review* 103.3, pp. 387–406.

Henry, Iain D (2022). *Reliability and Alliance Interdependence: The United States and Its Allies in Asia, 1949–1969*. Cornell University Press.

Hermann, Charles F (1969). "International crisis as a situational variable". In: *International politics and foreign policy* 2, pp. 409–421.

— (1972). "International crises: Insights from behavioral research". In: *(No Title)*.

Hogg, Tad and Lada Adamic (2004). "Enhancing reputation mechanisms via online social networks". In: *Proceedings of the 5th ACM Conference on Electronic Commerce*, pp. 236–237.

Hopf, Ted (1994). *Peripheral visions: Deterrence theory and American foreign policy in the third world, 1965-1990*. University of Michigan Press.

Hung, Angela A and Charles R Plott (2001). "Information cascades: Replication and an extension to majority rule and conformity-rewarding institutions". In: *American Economic Review* 91.5, pp. 1508–1520.

Hurwicz, Leonid (1946). "Theory of the Firm and of Investment". In: *Econometrica, Journal of the Econometric Society*, pp. 109–136.

Huth, Paul K (1988). "Extended deterrence and the prevention of war". In: *(No Title)*.

— (1997). "Reputations and deterrence: A theoretical and empirical assessment". In: *Security Studies* 7.1, pp. 72–99.

J. O'Gorman, Hubert (1986). "The discovery of pluralistic ignorance: An ironic lesson". In: *Journal of the History of the Behavioral Sciences* 22.4, pp. 333–347.

Jensen, Nathan M and Edmund J Malesky (2018). "Nonstate actors and compliance with international agreements: An empirical analysis of the OECD anti-bribery convention". In: *International Organization* 72.1, pp. 33–69.

Jervis, Robert (1989). *The logic of images in international relations*. Columbia University Press.

— (2017). *How statesmen think: The psychology of international politics*. Princeton University Press.

Kelman, Herbert C (1958). "Compliance, identification, and internalization three processes of attitude change". In: *Journal of conflict resolution* 2.1, pp. 51–60.

Kenny, David A (2004). "PERSON: A general model of interpersonal perception". In: *Personality and social psychology review* 8.3, pp. 265–280.

Keohane, Robert Owen and Joseph S Nye (2001). *Power and Interdependence. 4th*.

Kertzer, Joshua (2016). *Resolve in international politics*. Vol. 2. Princeton University Press.

Kertzer, Joshua D, Jonathan Renshon, and Keren Yarhi-Milo (2021). "How Do Observers Assess Resolve?" In: *British Journal of Political Science* 51.1, pp. 308–330.

Korte, Charles (1972). "Pluralistic ignorance about student radicalism". In: *Sociometry*, pp. 576–587.

Kreps, David M et al. (1982). "Rational cooperation in the finitely repeated prisoners' dilemma". In: *Journal of Economic theory* 27.2, pp. 245–252.

Kuran, Timur (1998). *Private truths, public lies: The social consequences of preference falsification*. Harvard University Press.

Kuran, Timur and Cass R Sunstein (1998). "Availability cascades and risk regulation". In: *Stan. L. Rev.* 51, p. 683.

Leeson, Peter T and Andrea M Dean (2009). "The democratic domino theory: An empirical investigation". In: *American Journal of Political Science* 53.3, pp. 533–551.

Li, Jian, Mauricio R Delgado, and Elizabeth A Phelps (2011). "How instructed knowledge modulates the neural systems of reward learning". In: *Proceedings of the National Academy of Sciences* 108.1, pp. 55–60.

List, John A (2006). "The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions". In: *Journal of political Economy* 114.1, pp. 1–37.

Lopez-de-Silanes, Florencio, Joseph A McCahery, and Paul C Pudschedl (2020). "ESG performance and disclosure: A cross-country analysis". In: *Singapore Journal of Legal Studies*, pp. 217–241.

Lu, Fangzhou, Lei Huang, and Sixuan Li (2022). "Sanctions and Social Capital: Evidence from the Russian Invasion of Ukraine". In: *Available at SSRN 4108129*.

Lupton, Danielle L (2020). *Reputation for resolve: How leaders signal determination in international politics*. Cornell University Press.

Mearsheimer, John J (2011). *Why leaders lie: The truth about lying in international politics*. Oxford University Press.

Mercer, Jonathan (2010). *Reputation and international politics*. Cornell University Press.

Miceli, Thomas J and Metin M Coşgel (1994). "Reputation and judicial decision-making". In: *Journal of Economic Behavior & Organization* 23.1, pp. 31–51.

Milinski, Manfred, Dirk Semmann, and Hans-Jürgen Krambeck (2002). "Reputation helps solve the 'tragedy of the commons'". In: *Nature* 415.6870, pp. 424–426.

Miller, Dale T and Leif D Nelson (2002). "Seeing approach motivation in the avoidance behavior of others: Implications for an understanding of pluralistic ignorance." In: *Journal of Personality and Social Psychology* 83.5, p. 1066.

Miller, Dale T and Deborah A Prentice (1994). "Collective errors and errors about the collective". In: *Personality and Social Psychology Bulletin* 20.5, pp. 541–550.

Miller, Gregory D (2004). *The shadow of the past: the influence of reputation on alliance choices*. The Ohio State University.

Mitchell, Sara McLaughlin and Elise Pizzi (2021). "Natural disasters, forced migration, and conflict: The importance of government policy responses". In: *International Studies Review* 23.3, pp. 580–604.

Morgan, T Clifton (1990). "Issue linkages in international crisis bargaining". In: *American Journal of Political Science*, pp. 311–333.

Nelson, Travis (2010). "When disaster strikes: on the relationship between natural disaster and interstate conflict". In: *Global Change, Peace & Security* 22.2, pp. 155–174.

O'GORMAN, HUBERT J (1975). "Pluralistic ignorance and white estimates of white support for racial segregation". In: *Public Opinion Quarterly* 39.3, pp. 313–330.

O'Gorman, Hubert J and Stephen L Garry (1976). "Pluralistic ignorance—A replication and extension". In: *Public Opinion Quarterly* 40.4, pp. 449–458.

Orme, John D (1992). *Deterrence, reputation and Cold-War cycles*. Springer.

Ostrom, Elinor and James Walker (2003). *Trust and reciprocity: Interdisciplinary lessons for experimental research*. Russell Sage Foundation.

Pearlman, Wendy (2018). "Moral identity and protest cascades in Syria". In: *British Journal of Political Science* 48.4, pp. 877–901.

Peterson, Paul E (1994). "The president's dominance in foreign policy making". In: *Political Science Quarterly* 109.2, pp. 215–234.

Peterson, Timothy M (2013). "Sending a message: The reputation effect of US sanction threat behavior". In: *International Studies Quarterly* 57.4, pp. 672–682.

Pevehouse, Jon C (2002a). "Democracy from the outside-in? International organizations and democratization". In: *International organization* 56.3, pp. 515–549.

— (2002b). "With a little help from my friends? Regional organizations and the consolidation of democracy". In: *American Journal of Political Science*, pp. 611–626.

Podolny, Joel M (2010). *Status signals: A sociological study of market competition*. Princeton University Press.

Press, Daryl Grayson (2005). *Calculating credibility: How leaders assess military threats*. Cornell University Press.

Raghunandan, Aneesh and Shiva Rajgopal (2022). "Do ESG funds make stakeholder-friendly investments?" In: *Review of Accounting Studies* 27.3, pp. 822–863.

Renshon, Jonathan, Allan Dafoe, and Paul Huth (2018). "Leader influence and reputation formation in world politics". In: *American Journal of Political Science* 62.2, pp. 325–339.

Renshon, Jonathan, Keren Yarhi-Milo, and Joshua D Kertzer (2023). "Democratic reputations in crises and war". In: *The Journal of Politics* 85.1, pp. 1–18.

Richard, Lebow (1981). *Between Peace and War: The Nature of International Crisis*.

ROLFE, MEREDITH (2012). "It Ain't What You Do, it's Who You Do It With: Distinguishing Reputation and Status". In: *The Oxford Handbook of Corporate Reputation*, p. 160.

Ross, Lee, David Greene, and Pamela House (1977). "The "false consensus effect": An egocentric bias in social perception and attribution processes". In: *Journal of experimental social psychology* 13.3, pp. 279–301.

Salganik, Matthew J, Peter Sheridan Dodds, and Duncan J Watts (2006). "Experimental study of inequality and unpredictability in an artificial cultural market". In: *science* 311.5762, pp. 854–856.

Salganik, Matthew J and Duncan J Watts (2008). "Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market". In: *Social psychology quarterly* 71.4, pp. 338–355.

— (2009). "Web-based experiments for the study of collective social dynamics in cultural markets". In: *Topics in cognitive science* 1.3, pp. 439–468.

Sartori, Anne E (2002). "The might of the pen: A reputational theory of communication in international disputes". In: *International Organization* 56.1, pp. 121–149.

— (2016). *International Reputation with Changing Resolve*. Tech. rep. Working Paper. Accessed March 2, 2021. https://asartori. mit. edu/sites ...

Schanck, Richard Louis (1932). "A study of a community and its groups and institutions conceived of as behaviors of individuals." In: *Psychological Monographs* 43.2, p. i.

Schelling, Thomas (1960). "The Strategy of Conflict (Cambridge, MA: Harvard UP)". In.

Schelling, Thomas C (1956). "An essay on bargaining". In: *The American Economic Review* 46.3, pp. 281–306.

Sechser, Todd S (2010). "Goliath's curse: Coercive threats and asymmetric power". In: *International Organization* 64.4, pp. 627–660.

Shamir, Jacob and Michal Shamir (1997). "Pluralistic ignorance across issues and over time: Information cues and biases". In: *Public Opinion Quarterly*, pp. 227–260.

Sharman, Jason C (2007). "Rationalist and constructivist perspectives on reputation". In: *Political Studies* 55.1, pp. 20–37.

Simmons, Beth A (2000). "International law and state behavior: Commitment and compliance in international monetary affairs". In: *American Political Science Review* 94.4, pp. 819–835.

Simmons, Beth A, Frank Dobbin, and Geoffrey Garrett (2006). "Introduction: The international diffusion of liberalism". In: *International organization* 60.4, pp. 781–810.

Snyder, Glenn Herald and Paul Diesing (2015). *Conflict among nations: Bargaining, decision making, and system structure in international crises*. Princeton University Press.

Sohn, YJ and Ruthann Weaver Lariscy (2014). "Understanding reputational crisis: Definition, properties, and consequences". In: *Journal of Public Relations Research* 26.1, pp. 23–43.

Solomon, Brittany C and Simine Vazire (2016). "Knowledge of identity and reputation: Do people have knowledge of others' perceptions?" In: *Journal of Personality and Social Psychology* 111.3, p. 341.

Sommerfeld, Ralf D et al. (2007). "Gossip as an alternative for direct observation in games of indirect reciprocity". In: *Proceedings of the national academy of sciences* 104.44, pp. 17435–17440.

Stimmer, Anette and Jess Gliserman (2022). "Disentangling norms, morality, and principles: the September 2019 Brexit rebellion". In: *European Journal of International Relations* 28.4, pp. 955–982.

Sunstein, Cass (2019). *Conformity: The Power of Social Influences*. New York University Press.

Sunstein, Cass R (1999). "The law of group polarization". In: *University of Chicago Law School, John M. Olin Law & Economics Working Paper* 91.

— (2001). *Designing democracy: What constitutions do*. Oxford University Press.

— (2002). "Conformity and dissent". In: *Available at SSRN 341880*.

— (2008). "Why groups go to extremes". In: *The Yale Law Journal* 110, p. 71.

— (2019). *How change happens*. Mit Press.

Tang, Shiping (2005). "Reputation, cult of reputation, and international conflict". In: *Security Studies* 14.1, pp. 34–62.

Tennie, Claudio, Uta Frith, and Chris D Frith (2010). "Reputation management in the age of the world-wide web". In: *Trends in cognitive sciences* 14.11, pp. 482–488.

Thierer, Adam et al. (2015). "How the internet, the sharing economy, and reputational feedback mechanisms solve the lemons problem". In: *U. Miami l. Rev.* 70, p. 830.

Thompson, Larry Dean (2004). *Corporate Scandals, why they happened and why they may not happen again*. Chautauqua Institution.

Thorsteinson, Todd J and William K Balzer (1999). "Effects of coworker information on perceptions and ratings of performance". In: *Journal of Organizational Behavior* 20.7, pp. 1157–1173.

Thrall, Calvin (2021). "Public-private governance initiatives and corporate responses to stakeholder complaints". In: *International Organization* 75.3, pp. 803–836.

Tingley, Dustin H and Barbara F Walter (2011). "The effect of repeated play on reputation building: an experimental approach". In: *International Organization* 65.2, pp. 343–365.

Tirole, Jean (1996). "A theory of collective reputations (with applications to the persistence of corruption and to firm quality)". In: *The Review of Economic Studies* 63.1, pp. 1–22.

Tomz, Michael (2007). "Domestic audience costs in international relations: An experimental approach". In: *International Organization* 61.4, pp. 821–840.

— (2012). *Reputation and international cooperation: sovereign debt across three centuries*. Princeton University Press.

Treisman, Daniel (2004). "Rational appeasement". In: *International Organization* 58.2, pp. 345–373.

Tversky, Amos and Daniel Kahneman (1974). "Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty." In: *science* 185.4157, pp. 1124–1131.

Walter, Barbara F (2009). *Reputation and civil war: Why separatist conflicts are so violent*. Cambridge University Press.

Wapner, Paul Kevin (1996). *Environmental activism and world civic politics*. State University of New York Press.

Weber, Max (1978). "Economy and society: An outline of interpretive sociology. Berkeley". In: *University of California Press* 2, pp. 243–245.

Weisiger, Alex and Keren Yarhi-Milo (2015). "Revisiting reputation: How past actions matter in international politics". In: *International Organization* 69.2, pp. 473–495.

Willinger, Marc and Anthony Ziegelmeyer (1998). "Are more informed agents able to shatter information cascades in the lab?" In: *The economics of networks: Interaction and behaviours*. Springer, pp. 291–305.

Wilson, Robert (1985). "Reputations in games and markets". In: *Game-theoretic models of bargaining* 96, pp. 27–62.

Wohlforth, William C et al. (2018). "Moral authority and status in International Relations: Good states and the social dimension of status seeking". In: *Review of International Studies* 44.3, pp. 526–546.

Wolford, Scott (2007). "The turnover trap: New leaders, reputation, and international conflict". In: *American Journal of Political Science* 51.4, pp. 772–788.

Yarhi-Milo, Keren (2018). *Who fights for reputation: The psychology of leaders in international conflict*. Vol. 156. Princeton University Press.

Zarolia, Pareezad, Max Weisbuch, and Kateri McRae (2017). "Influence of indirect information on interpersonal trust despite direct information." In: *Journal of personality and social psychology* 112.1, p. 39.

# Reputation Between Disciplines:

# A Group-Based Theory of Firm Reputation in IR

Hayley Pring

Feb 2025

## 1 Introduction

In the 20th century a prevailing view was that the relationship between corporations and society was governed by a coherent regulatory environment, where political actors and institutions were responsible for addressing problems or providing public goods. To the extent that corporations did enter the political sphere, it was primarily a means of securing their own economic interests, often by engaging in activities like lobbying for favorable policies. In this framework, only governments and state institutions are political and responsible for social welfare, while firms remain as purely economic actors (Sundaram and Inkpen 2004). The purpose of the firm is to increase its market value (M. C. Jensen 2002). The pursuit of activities under the umbrella of corporate social responsibility ultimately diverts profits (Henderson 2001), and is seen as encompassing "activities that used to be responsibilities of democratically elected and legitimized government" (Matten 2009, p. 570).

In the 21st century, however, the revenue of several multinational corpora-

tions eclipses that of some nation-states. In 2024, Microsoft's total market value was greater than that of Russia, South Korea, Spain, the Netherlands, Canada, Mexico, Brazil, Australia and Indonesia.[1] Large public companies like Apple, Saudi Aramco, Amazon, Alphabet, Meta, Alibaba Group, Samsung, JPMorgan Chase, and Walmart rival the GDP of several countries (Rescue 2024). Within states, the business sector accounts for an overwhelming majority of the value of GDP; within the OECD this figure is 72 percent (Manyika and Birshan 2021). Multinational corporations can and have hired private security firms or private military contractors to protect business operations and assets (Avant 2004). They spend billions in lobbying with unrivaled access to elected representatives across countries, and coordinate complex in-house public policy agencies. In some instances where there is low governance or high conflict, corporations have acted as de facto governments in the provision of public goods and infrastructure (Bradly and Nathan 2019; Backer 2015). For scholars of political science and international relations the so-called "goldilocks balance" between the power of firms and the power of the state has titled (Zingales 2017, p. 125). As firms gain increasing power, traditional definitions that once separated governments from corporations—such as the legitimate use of force (Waltz 1979, pp. 103–104)[2] and the role of law and institutions in maintaining order—are becoming increasingly blurred. This is especially true of multi-national corporations (MNCs), whose operations extend beyond the limits of any single state.

Large corporations wield enormous influence over traditional cross-border political issues, complicating, and in some instances undermining, cooperation between states over international governance issues. This evolution of the firm goes beyond a simple definition of a "nexus of contracts" (Meckling and M. C.

[1] GDP and CMC are not equivalents; CMC is a measure of the total value of publicly listed company shared, and GDP the total value created by goods and services in a country.

[2] Notably, Waltz includes the use of economic sanctions in his definition of the use of force.

Jensen 1976) where corporations are thought to affect only contracting parties. Instead, MNCs represents a significant concentration of economic and political power in the hands of unaccountable individuals with, at best, limited responsibility to shareholders (and even then, only if the firm is publicly listed). This is not a new phenomenon, but rather an idea that has been percolating for a century, with references to the uneasy domination of the firm over politics present in the early 20th century. Consider the iconic work of Berle and Means (1933) in *The Modern Corporation and Private Property*:

> "The rise of the modern corporation has brought a concentration of economic power which can compete on equal terms with the modern state—economic power versus political power, each strong in its own field. The state seeks in some aspects to regulate the corporation, while the corporation, steadily becoming more powerful, makes every effort to avoid such regulation. ... The future may see the economic organism, now typified by the corporation, not only on an equal plane with the state, but possibly even superseding it as the dominant form of social organization".[3].

In theory these companies can be regulated and reined in by the state. Yet, companies avoid regulation by lobbying elected representatives to kill or water down unfavorable legislation,[4] by directly lobbying the regulator to avoid enforcement (Lambert 2019), or by using their vast legal resources to conceal their behavior to avoid detection. Take the case of Dupont who, for instance,

[3]Quoted from: (Zingales 2017)

[4]Beyond the firm-level benefits, the internal incentive-structure of the firm rewards this behavior, with Brown and Huang (2020) finding the share price of firm executives who visited the white-house between 2009-2015 increased by 1 percent in the two months following the visit. As Zingales (2017) points out, this number is small, "until you discover that during Obama's presidency, the chairman and chief executive officer of Honeywell international visited the White House 30 times, while the head of General Electric visited 22 times."

successfully hid the company's contamination of a West Virginia water supply for three decades under trade secret law (Shapira and Zingales 2017). Unlike elected governments, citizens do not always have the same rights over companies to know the impact of their activities on public goods, or make freedom of information requests to increase their awareness of loopholes or inefficiencies in regulations (Horwitt 2022; M. Dean 2021). This effectively weakens the counterbalance of firm lobbying: citizen pressure on elected representatives to legislate or enforce regulations.[5] Even taxation has proved to be a difficult regulatory task. It was not until 2021 that nearly 140 countries reached an agreement on a 15% global minimum tax rate for multinational corporations—an agreement the United States later withdrew from through a presidential executive order in early 2025.

Corporations shape and dominate domestic politics, largely unchecked by market pressures beyond investor interests.[6] However, the areas they influence - whether deliberately or as a byproduct of profit-seeking behavior - often intersect with governance challenges that states already struggle to cooperate over due to conflicting interests, high uncertainty, and incentives to defect. An easy example is the environment and more specifically climate change, where just 100 corporations are responsible for over 70% of the world's greenhouse gas emissions since 1988 (Riley 2017).

Thus the role of the firm in international relations becomes not only more relevant, but dominant, in several cooperation issues previously faced by states alone. But if firms, like states, face conditions resembling those of anarchy, where enforcement mechanisms by state regulations (through will or complex-

[5]This balance has grown starker in some jurisdictions like the US, where restrictions on corporate donations was unleashed following the Citizens United ruling in 2010.

[6]See Zingales (2017) for criticism of investor restraint over MNCs.

ity) are weak, alternative means are needed to either bring firms into compliance or encourage them to make and keep agreements. How can firm cooperation or self-regulation in world politics emerge? One mechanism that exists to restrain or regulate these actors is reputation.

One of the few areas in IR where some consensus exists is that tapping into states' concerns for their reputation is one of the most effective methods of increasing levels of compliance and cooperative behavior. As Guzman (2002, p. 1849) suggested, "In the absence of other enforcement mechanisms, then, a state's commitment is only as strong as its reputation. When entering into an international commitment, a country offers its reputation for living up to its commitments as a form of collateral". But the origins of these ideas came from economics. Since many of the foundations of IR theory were built on scholars analogizing the behavior of states to that of firms (Waltz 1979), it's unsurprising that reputation traveled from one discipline to the other. That being said, the traditional focus of realist thought in IR on states as the major actors skewed the focus of research on reputation to states in the domain of conflict. Certain types of reputations, like resolve, were prioritized since these reputations were pivotal to a states' survival in an anarchical system.

Microeconomics and management studies developed a parallel study of reputation in the context of firm behavior in the market. While the two studies of reputation overlap in many respects, particularly with their concern for the history-dependent nature of reputation, a split between the two emerged as a result of the institutional and legal structures of the market versus the international system, the relevant audiences, and the stakes of the issues. Given this divergence and the evolving role of firms in IR studies, the question arises as to

which discipline provides a stronger framework for explaining firm behavior.

This paper uses insights from the different disciplines and offers a theoretical re-examination of the study of reputation in international relations, arguing that concern for reputation influences firm behavior, including cooperation and self-regulation. Unsurprisingly, since the original strains of realism have their roots in micro-economics and anarchy as a type of market, the two disciplines share similar notions of reputation, remaining fairly loyal to Schelling's (1960) ideas on costly signaling and deterrence. The IR model of reputation that relies on a history-dependent definition of reputation, treats observers as largely homogeneous,[7], and focuses on resolve as the major reputational type, has found many proponents as well as critics.

However, the logic of why reputation has developed with these major assumptions in IR, and whether this model captures the logic of firm behavior has not been subject to rigorous theoretical or empirical scrutiny. This paper analyzes the different perspectives on reputation in both disciplines, noting where they begin to diverge and why some of the assumptions in IR models of resolve are better suited to explaining state-based actors and reputations for resolve,[8] but fall short in explaining firm behavior in world politics. I argue that while both disciplines have sought to explain the role of their primary actors within their respective spheres - markets versus world politics - understanding firm reputation in IR requires a new approach that integrates insights from both. The aim of this paper is to present a richer theoretical model that integrates both perspectives given economics-expertise on the firm, and IR's expertise on

[7]Exceptions include J. Kertzer (2016) but as I detail below, there is more work to be accomplished on heterogeneous audiences.

[8]Renshon et al. (2018) showed that 69% of IR studies on reputation focused on the state, 21% of studies focused on leaders, and only 10% of studies clustered around the category of 'other'.

the international system.

Rather than dismissing either discipline's approach, I offer an alternative framework that begins with the observations that stakeholders belong to heterogeneous groups whose vantage points and interactions with the firm structures their expectations of the firm. Multiple audiences allows for the logic of multiple reputational types, and creates a strategic situation where firms must signal their type to different observers in order to maintain a competitive advantage over their competitors or new entrants. I specify how an exogenous shock or crises can prompt one or more groups of observers to desire a newly salient reputational type from firms.

While audiences can use history to determine type for highly salient primary reputations - such as resolve or quality - where the audience has greater awareness of these reputations and firms are consistently signaling type, observers have less information about secondary reputations or simply lack recall. Under these conditions, observers instead compare firms within their collective 'groups' based on identifiable criteria to gauge the firm's type. Importantly, when one member of this group adjusts their behavior, they reset the reputational benchmark of the group. This process is dynamic and iterative; what constitutes a 'good' or 'bad' type in a given situation is constantly adjusted as member's of the group move and observer's update their beliefs.

Moreover, it's not *just* consumers updating beliefs; if reputation is a positional good, then competitors within the group are aware that the movement of their peers can have a reputation-stealing effect on their own reputation. The firms within the group update their beliefs about how they are being perceived.

This group-based model of reputation thus induces a cascade effect, where the actions of an early mover forces other members of the group to conform. The effect, however, is temporally bound - if the salience of the situation declines, and observer attention to this reputation type wanes, firm signaling drops off.

This paper proceeds with a discussion of the role of reputation for firms in management and micro-economics, and traditional approaches to reputation in international relations. I then outline the three key premises of reputational models: reputational type, audience homogeneity, and history-dependence. I then offer a competing group-based theory of reputation, and finally suggest some testable implications of the theory.

In this paper I seek to make three contributions. First, I offer an overarching conceptual framework of reputation to better understand the group-dynamics that shape both actors' concern for reputation and how observers assess reputational type, as an alternative model to history-dependent reputations. I maintain that while historical behavior is still important for reputations to form, if observer knowledge is low as a function of a less-salient reputation, then observer's will use heuristics and easily accessible information to make judgments on type; the actor's peers. Second, I hope to explain the divergence of reputation between disciplines that had similar starting points on reputation, and integrate insights from both bodies of research. One reason why the economics and management literature, unlike IR, offers insights into group-behavior is because scholars from this field have always acknowledged the existence of different stakeholders for firms. When different groups of observers exist whose relationship with the firm structures both their expectations of the firm, as well as the degree of informational asymmetries, different reputational types and ways of

signaling naturally arise. The focus in IR, with few exceptions, on resolve as the major reputational type, states as the major actors, and unitary audiences, has narrowed down expectations of observers and actors operating in silos. Lastly, I introduce pathways through which reputation effects firm behavior as a result of group-dynamics; cascades, early-movers, and dueling audiences.

# 2 Reputation between Disciplines

## 2.1 Defining Reputation

In the management and economics literature, reputation is conceptualized in different ways among sub-disciplines. For instance, economists view reputations as functional: they are either traits or signals (Stigler 1962; Weigelt and Camerer 1988), where informational asymmetries force observers to rely on proxies to determine type. Corporations utilize "strategic projection" (advertising, campaigns, conference calls) to signal hidden qualities and win over observers (consumers, investors, employees) (V. P. Rindova and C. J. Fombrun 1999; Shapiro 1983). For these scholars, it's important to note that reputations are the result of prior resource allocation - i.e. firm investments in product quality (Myers and Majluf 1984). In contrast, in the strategy literature, reputation is an intangible economic asset of the firm that confers on them a competitive advantage in the marketplace (Milgrom and J. Roberts 1982; Hall 2009; C. J. Fombrun 2005; G. R. Dowling 2004). The functional utility of reputation is "distinguishing the organization from its peers" (King and Whetten 2008, pp. 192–3). Reputation also acts as a barrier to mobility (Caves and Porter 1977).[9] This is because reputations are thought to derive from unique internal features of the firm (Dutton and Dukerich 1991) but are externally perceived. Consistent with a view of

[9]If reputations take time to coalesce in audience minds, and building good reputations is costly and difficult for rivals to replicate, reputation creates barriers to market-entry for new entrants or competitors (Hall 2009; P. W. Roberts and G. R. Dowling 2002).

reputation as a resource, a firm's reputation is the outcome of a competitive game "in which a firm signals its key characteristics to constituents in order to maximize its economic and non-economic status" (Walsh and Beatty 2007, p. 128). Marketing scholars hold similar views, where reputations are useful - or functional - insofar as they stabilize firm revenue and induce repeat purchases/build customer loyalty (C. Fombrun and Shanley 1990; Aaker 2012).[10]

In contrast, definitions of reputation in IR are plentiful and the differences between them are not just semantics. Mercer (2010, p. 6) offers one variation of this, "Reputation is a judgment of someone's character (or disposition) that is then used to predict or explain future behavior". This tracks with similar definitions, such as Dafoe, who defines it as, "Reputation refers to beliefs about an actor's persistent characteristics or behavioral tendencies based on the latter's past behavior" (Dafoe et al. 2014). Renshon et al. (2023) uses a more socially-conscious definition of reputation as "socially shared beliefs about an actor's behavioral tendencies", and Yarhi-Milo (2018) opts for a vaguer definition that divorces reputation from the necessity of historical behavior, here reputation is, "the belief that others hold about a particular actor". The variation in whether reputation requires observation of past actions is clear in these three definitions.

The defining characteristic of economics and management definitions of reputation is that signaling reputational type is a functional means for a firm to gain competitive advantage over their competitors and differentiating the firm from others, rather than just a tool for observers to assess how they interact with the actor. Both disciplinary definitions are similar in their concern for historically-derived reputations, although this quality varies in explicit defini-

[10]There are additional sub-disciplinary views in the organizational behavior and organizational psychology literature that defines reputation as the shared cultures and identities of employees and managers (Albert and Whetten 1985).

tions.

# 3 Signaling and attributes

### 3.0.1 Economics

While Schelling published in the economics discipline, micro-economics and the theory firm reputation took a different path with Spence (1974) work on signaling.[11] Spence does not mention reputation, but his model that strategic actors deliberately take actions that send signals of information about unobserved or hidden attributes to influence their counterparts' beliefs was the basis of Kreps and R. Wilson (1982) and Milgrom and J. Roberts (1982) foundational work on how market actors facing repeated competitive interactions make decisions under conditions of informational asymmetry. Milgrom and J. Roberts (1982) predation model has similar assumptions and implications of T. C. Schelling (1967) and the later work of Fearon (1995) and Fearon (1994) on signaling resolve and rational reasons for states to go to war, even when it is costly. In Milgrom and J. Roberts (1982), the authors respond to the prevailing position in economics that predation is an irrational strategy for attempting to gain or maintain a monopolistic position because predation is costly to the incumbent and unlikely to work as a successful strategy in driving out rivals. This is because competitors understand that price-cutting measures can only be temporary. Furthermore, even if the predatory monopolistic firm is successful in chasing away the rival, any subsequent attempt to raise prices to reap the benefits of the established monopoly position would serve to attract new entrants. Thus, any monopoly gains of predation are short-lived and the costs outweigh the benefits. Milgrom and J. Roberts (1982) reject this outcome, and instead argue that predation

[11] Though Spence 1974 cites Schelling for his signaling model on the "incentive structures and outcomes of n-person symmetric, binary choice situations"

can be a rational, successful, and a profit-maximizing strategy, not because of the short-term gains of driving out any specific rival, but because engaging in predation helps the firm establish a reputation as aggressive. New entrants learn of this reputation, and are deterred from entering the market since they expect the incumbent firm to engage in predatory behavior. This is still similar to the function of reputation in the conflict signaling model albeit in a different context, take for instance the following extract from T. C. Schelling (1967):

> "It is often argued that "face" is a frivolous asset to preserve, and that it is a sign of immaturity that a government can't swallow its pride and lose face ... But there is also the more serious kind of "face," the kind that in modern jargon is known as a country's "image," consisting of other countries' beliefs (their leaders' beliefs, that is) about how the country can be expected to behave. It relates not to a country's "worth" or "status" or even "honor," but to its reputation for action. If the question is raised whether this kind of "face" is worth fighting over, the answer is that this kind of face is one of the few things worth fighting over. Few parts of the world are intrinsically worth the risk of serious war by themselves, especially when taken slice by slice, but defending them or running risks to protect them may preserve one's commitments for actions in other parts of the world at later times. "Face" is merely the interdependence of a country's commitments; it is a country's reputation for action, the expectations other countries have about its behavior."

The similarities are clear: risking war is not a beneficial strategy with any particular state, but it is ultimately less costly to deter future aggressions by making war with that state less attractive in developing a reputation for resolve. Both

models solve the paradox of war and the chain-store paradox by acknowledging asymmetric or incomplete information: new entrants are uncertain that incumbent firms will rationally choose between predation and peaceful coexistence, and aggressive states are uncertain that target states will rationally choose between going to war or backing down. It's important to highlight the necessity of incomplete information to all of the foundational models of when reputation matters.

What's interesting is that in early models of reputation in IR, the domestic and the international were treated distinctly; the audience for the state was other states. However, because Milgrom and J. Roberts (1982) are modeling reputation in the context of the market, a few things are possible and distinct from international politics. One of which is that two audiences can effect the strategic behavior of the monopolistic firm. First, the incentive not to engage in predatory behavior is the consideration that cost-cutting is an unsustainable strategy and eventually the monopolistic firm would raise prices once the new entrant left in order to reap the benefits of their monopoly position in the market. The assumption is that a reputation for predation only has valuable currency among potential new entrant firms, not consumers. For consumers, the reputation for quality (and price) matters more, and is potentially in conflict with the reputation for predation among firms. Later models of reputation and resolve in IR would consider a similar idea that signaling conflict has a double-barreled treatment (J. D. Kertzer and Brutger 2016).

### 3.0.2 IR

Schelling's pioneering research on strategic behavior set the stage for IR's research traditions in reputation and conflict. While his often cited work heralds from his 1996 book *Arms and Influence*, his first iterations of reputation orig-

inated in a 1956 article, *"An Essay on Bargaining"* (T. C. Schelling 1956). In that paper Schelling suggested that the processing of decomposing threats into a series of steps would help an actor signal their reputational type.[12]

Schelling was suggesting that observers make inferences based on behavior about an actor that are informative and can help classify how and which actors may make decisions now or in the future. Jervis (1989) *Logic of Images* extended Schelling's position to argue that behaviors that discriminate between possibilities help observers make decisions about how to engage with, react to, or perceive a threat. As highlighted in the *Logic of Images*, the process of making inferences is complicated by the ability of actor's who are incentivize to project an image of themselves that may lead observer's into making inferences about the actor's type that may not be true.[13] However, Jervis's definition of signaling reputation as being the point in which an actor follows through on what they said they would do aligns more closely with reputations for resolve, where resolve refers to maintaining a policy or stance despite temptations or motives to back down.

Instead, actions actors take that reaffirm the 'type' they attempt to project is a more befitting outline of signaling reputation. What Jervis is getting at when he outlines reputational signaling is credibility. This is reinforced when he states, "Thus I might be believed to be ready to carry out my promises

[12] "To make the threatened acts divisible, the acts themselves may have to be modified. Parts of an act that cannot be decomposed may have to be left out; ancillary acts that go with the event, though of no interest in themselves, may be objects to which a threat can effectively be attached. For example, actions that are only preparatory to the main act, and by themselves do no damage, may be susceptible of chronological division and thus be effective objects of the threat. The man who would kick a dog should be threatened with modest punishment for each step toward the dog, even though his proximity is of no interest in itself" (T. C. Schelling 1956, p. 298).

[13] The deception that Jervis discusses about states in world politics is, in someways, more true of non-state actors who may be responding to multiple audiences than that of states.

but not my threats; to send valid Signals about my capabilities but not about my intentions; to be willing and able to keep my word in the economic area but not on military issues."(Jervis 2017a, p. 119). Jervis is suggesting actors can send signals on many different reputational types, however the carrying through, keeping their word, and validity are all references to credibility, which is different to a reputational type more generally. After all, actors can develop reputations of being fickle or non-credible in one area, but not another.[14]

How observers assess reputational type is complicated by the large number of indices and, in some cases, countervailing characteristics that reveal information about an actor. Unsurprisingly most of the theoretical frameworks on understanding how observers assess reputational type are clustered around resolve, and therefore involve the role of past behavior (Jervis et al. 1985; Yarhi-Milo 2018; Weisiger and Yarhi-Milo 2015; Mercer 2010; T. C. Schelling 1956; T. Schelling 1960; T. C. Schelling 1967; J. Kertzer 2016; J. D. Kertzer et al. 2019), the costliness of the signal (Fearon 1994; Fearon 1995), current capabilities (Press 2005), current interests (Danilovic et al. 2020; Diehl 1992), and then leader-level attributes (Horowitz and Stam 2014; Gelpi and Grieco 2001). Given the strategic environment in which actors operate, more recent work has shed light on how observers confront difficult decisions of adjudicating between signals. For instance, J. D. Kertzer et al. (2021) use a conjoint experimental framework to show that the American public use a combination of signals about an actor's capabilities, stakes, mobilization and past behavior to make judgments about an actor's type to predict future behavior.

[14]Issue-linkage - where states can enforce an agreement by linking it to anther issue, or punish a state for cheating in one area in different domain - is different from the idea of an actor building the same reputational type across different contexts if they are not behaving consistently.

While this work reinforced the notion that "past actions" amount to a running tally of how reputations are witnessed by past observers, there are issues in using conjoint or vignette experimental designs to assess whether observers place weight on historical behavior to determine reputational type. These studies take place in an experimental setting of prompting the participant to recall an instance of past behavior, where the actor either did or did not follow through (Tomz 2007; Renshon et al. 2018; Casler et al. 2023; J. D. Kertzer et al. 2021; Tingley and Walter 2011). For example, in J. Kertzer (2016) and Brutger and J. D. Kertzer (2018), respondents are given treatments about past actions. Similarly, in Goldfien et al. (2023), the authors prompt respondents to think about a leader's domestic repression from "a few years ago" as a treatment, as well as Arcadia's previous foreign policy fighting.

The problem is that these models build in the assumption that observers' would be able to recall this information, and therefore this attribute is *actually* used by observers to indicate type. However, none of these studies have tested whether observers are aware of past behavior. The reason why observers place greater weight on current actions over past behavior (Crescenzi 2018; Press 2005) is because recency bias makes this information more cognitively available for respondents. Although political scientists are increasingly acknowledging the role of heuristic short-cuts in making judgments (Gigerenzer and Gaissmaier 2011), a fact that seems to be forgotten is that this phenomenon was introduced to the discipline by Kahneman and Tversky (1972) and Kahneman et al. (1982) who explicitly state that availability is a function of recall or associative distance. Recent information weighs more heavily than historical information because it's easier to recall. Simply put, by giving individuals immediate information about both recent behavior and past behavior in an experimental setting, the effect of

testing recall is undermined – you are supplying the respondent with information on behavior that they will weigh equally (in terms of availability), effectively making history seem more, or equally, important for the observer than it may be.

For most political science research, the gap between the lab and the real world is challenging but the former is justified as a necessary step to overcoming the fundamental problem of causal inference. However, in the specific case of reputation where informational asymmetries and uncertainty of type are conditions for its existence, conjoint and vignette experiments that supply respondents with information eliminates the role of availability as the mediating variable that explains why some indicators receive more weight in assessing type. However, these criticisms perhaps apply more generally to observers of a less known type. Under conditions of hyper-aware participants for a highly salient reputational type, where past incidents were universally known or relatively recent, then the assumption that respondents recall history is less problematic.

The role of past behavior, or a vertical comparison of behavior,[15] in influencing how observers assess reputation likely does hold for prominent states or certain leaders, where either the time horizon is not too long ago in the case of the latter, or the event is highly salient in the case of the former. However, when we introduce a large universe of actors (firms), operating in different sectors, where audiences (consumers) may not be exclusively interacting with that single sector or firm, awareness of past behavior is necessarily lower. Under these conditions, situational elements and horizontal comparisons between alike actors are more likely reference points for observers to make judgments about reputation. It is in this context of many competing actors at once that the

[15] I use vertical and horizontal comparisons of behavior to refer to reputations that are assessed over time (vertically, from past to present) versus reputations that are assess by comparing peers in the present (horizontally across peers.

ability for individuals to recall specific incidents of past behavior for each firm, where the past event was either highly salient or occurred in short succession, seems extremely unlikely. The deficit in the IR literature on how the conditions under which observers form ideas of reputation vary by actor and reputational type has largely been the result of the continued focus on states as the units of observation, or the treatment of leaders as developing the same reputations.

# 4 Reputational Types and The Nature of the Audience

For many years, a broad consensus across different strands of economics accepted that reputation referred to stakeholders' interpretation of 'what a firm stands for', assuming that all stakeholders shared the same values and that customers, investors, employees, and competing firms had uniform expectations of firm behavior. The focus was on a group of heterogeneous actors who sent signals to a homogeneous audience, who received the signal and made a decision (Spence 1974). However these definitions were challenged by a competing position that viewed stakeholders as distinctive observer groups, based on their own needs (C. J. Fombrun 1996), and their particular relationship with the firm (Walsh and Beatty 2007; G. Dowling 2000). The two core findings were that (1) observer vantage point (i.e. their relationship to firm) structures observer expectations of the firm, and (2) firm's engage many stakeholders (or types of observers). The logical implication is that firms can hold multiple reputations, as different observers interpret signals in varying ways (Wartick 2002; Walsh and Beatty 2007).

If we take these insights and refer back to the discussion above about signaling, the effectiveness of a market signal is contingent on audiences in two key ways. First, a signal can only be effective at reducing uncertainty of type to the extent that the audience that *would* use this signal is paying attention (Pollock and Gulati 2007; Connelly et al. 2011; Madsen and Rodgers 2015, p. 340). Second, the effectiveness of market signals depends on interpretation (Heil and Robertson 1991). A body of work in management and economics emerged to consider when signals mattered (M. Jensen and Roy 2008; Podolny 1994; Spence 1978), what signals mattered (Pollock and Gulati 2007; Podolny 1993; Kim and M. Jensen 2014; Negro et al. 2015) and ultimately why heterogeneous audiences helps explain differences in signal evaluation (Falchetti et al. 2022; Sharkey et al. 2023; Durand and Haans 2022).

In contrast, in IR perception and signaling were not studied together. This was remarked on by Jervis (2017b, p. xvi) who commented in a footnote of the 2017 version of *Perception and Misperception in International Politics*, "in a better world, signaling and perception would be not only discussed between the same two covers, but also intimately tied together. Unfortunately this was not the case in my own work, and in subsequent years the two literatures have continued on diverging paths, to the detriment of both." This issue is apparent in a number of studies, which paid little attention to individual-level characteristics among the audience (Horowitz and Stam 2014; Davies and Johns 2013; Tomz 2007; Levendusky and Horowitz 2012).[16] Recent work in the audience cost literature has started to push against this idea, introducing the notion that i) observers assess reputational types differently, and ii) decomposing audiences

[16]homogeneity assumption in the treatment of reputation in two ways. Huth (1997, p. 79) has argued that reputational theories in IR assume, "1) The cognitive processes by which reputational inferences are drawn are similar across individuals; and 2) that the content of whatever inferences are drawn are also similar across individuals".

leads to insights on why some observer value some signals more than others. One such work that considers the role of multiple audiences is Clare and Danilovic (2010). These authors take aim at T. Schelling (1960)'s 'indiscriminate reputation' building and multiple audiences. Their study suggests that the presence of multiple rivals (as audiences) incentivizes states to signal reputation. However, the authors do not consider the idea that audiences may have different preferences over reputational types, that audiences may be located at different vantage points across time and space, such that it affects how they perceive indices and signals of behavior.

More recent and innovative work by J. D. Kertzer and Brutger (2016) argues that there is no unitary logic of the audience, different types of people have different preferences over reputational type and propensities to impose reputational penalties, depending on their dispositional traits. The authors focus on individual-level variation in international trust, nationalism, military assertiveness, and political ideology, and find that hawks and doves vary in the type of reputation they care about, and how they receive signals related to consistency and belligerence (J. D. Kertzer and Brutger 2016). This research has been helpful in understanding the relationship between actors who build reputations, and the audiences that observe them. However, because J. D. Kertzer and Brutger (2016) are using the level of analysis as an elected state leader with the focus of observation on domestic audience costs, in some ways the only relevant audience is the domestic public. Therefore, for J. D. Kertzer and Brutger (2016) audience heterogeneity is modeled off of policy preferences as a function of dispositional traits and attitudes, without considering whether the relationship between the audience and the actor helps structure the audience's preferences over the actor's type. This makes sense in the context of J. D. Kertzer and Brutger (2016)

work, however when we replace leaders with firms as the object of analysis, we introduce a multiplicity of actors, and in addition to consumers, a number of other stakeholders whose expectations of the firm are structured by their engagement with the firm not just their individual disposition in isolation from the firm.

It's here where the management and economics literature offers meaningful insights into how considering audience heterogeneity by groups rather than internal disposition can help explain how concern for reputation affects firm behavior. We can consider heterogeneous audiences in this context as the division of observers into subgroups characterized by the specific engagement type and engagement frequency of those observers with the actor, who possesses different levels of knowledge, preferences, and sanctioning abilities. To be clear, the utility of these insights extends beyond the firm-level to shedding light on the gap between public audiences and elite audiences, whose interactions with leaders and states differs (J. D. Kertzer et al. 2019; J. D. Kertzer 2022; Jost et al. 2024), as well as between types of foreign publics (Bell and Quek 2018).

# 5 History Dependence of Reputation

The above discussion on signaling and audience type brings to attention the issue of whether reputation is always history-dependent. The dominant assumption in IR is that reputation is a history-dependent concept. For both the literature on reputations for resolve and cooperation, the premise that underpins these theories of behavior is that commitments are interdependent; this might be across issue area, regions, and actors (Weisiger and Yarhi-Milo 2015).[17] In some ver-

[17]Reputation being conceptualized as a historically-built phenomenon stems from rationalist understandings of IR. While rationalist definitions classify reputation as "the degree to which

sions of this theory, a state will observe how state A behaves towards state B in a given dispute, and use this information as a template to anticipate behavior in a different dispute, in a separate arena. For instance, Jervis (2017a, p. 85) claims that, "Past behavior can be used as an index to the state's characteristics and how it will behave in the future in a way that does not depend on signals". Empirical examples of this include Treisman (2004) who illustrates the logic of Kreps et al. (1982) 'chain store paradox' model in IR by showing how states' invest in reputations for appeasement which seemingly slowed the disintegration of the British Empire. In contrast, Press (2005) finds little evidence that German or US policy-makers were able to infer behavior of each other based on past records of resolve. Press argues temporal dynamics mean governments only pay attention to situational attributes over dispositional attributes during crisis. His argument is that, contrary to conventional wisdom, credibility is not related to a state's "history of keeping or breaking commitments" (Press 2005, p. 143), but observers believing "threats and promises that are backed by sufficient power and serve clear interests" (Press 2005, pp. 8–9). A big issue for this argument is that the audience of observers are a smaller pool of experienced decision-makers with information on the state's past behavior; the idea that all states - or indeed all actors - have a blank canvas in every crisis is difficult to maintain. For both Press (2005) and Mercer (2010), the traditional realist variables of current, material interest and power are what explains outcomes in crisis.

Yet there is by no means a consensus on the role of history in the literature on reputation. Empirical work examining whether historical behavior did build

an actor reliably upholds its commitments, based on a record of that actor's past behavior". In contrast, constructivists differ by suggesting "reputation is the generally shared opinion of a referent's character or nature, based on a wide range of information, associations and social cues" (Sharman 2007, p. 20).

reputations that influenced policymakers has found limited evidence (Hopf 1994; Richard 1981; Snyder and Diesing 2015). Wolford (2007) argued that new leaders face incentives to build reputations for resolve in the shadow of leadership turnover, while antagonists are not deterred by a state's past resolve, since new incumbents may possess differing levels of resolve. The result is that past behavior is irrelevant, and adversaries will seek to test new incumbents. Challenging the idea of T. Schelling (1960) on indiscriminate reputation building across time and space are Snyder and Diesing (2015), Jervis (1989), and Huth (1997). For Huth (1988), historical behavior of a state only impacted current behavior to the extent that it was in reference to the same adversary with whom it shared a history of conflict (Danilovic et al. 2020). Similarly, in order for historical behavior to matter for reputations for resolve as a deterrent, other scholars have found there must be a common variable (or perhaps, a common focal point). For instance, Danilovic (2002) found that when a trajectory of conflict occurs in the same region, past behavior of one member of a dyad can influence how the other member of the dyad responds.

There are two issues with the current debate about the relevance of history-dependence in reputation formation. The first is that scholars that are seeking to argue against game-theoretic models that past behavior does not impact current behavior do not have a suitable argument for explaining how reputations form in the absence of historical record. The assumption then, is that reputation in general neither matters as an independent variable or an outcome variable. If reputation does not matter, actors in world politics should not be motivated to create, maintain, or avoid specific types of reputations.

It might be the case that theoretical and empirical evidence for the role of

history in reputations only holds for some actors under some reputational types, such as resolve, not reputations in other lower-stake arenas. However, given the overwhelming focus on resolve - in some cases authors using the reputation as a shorthand for resolve (Dafoe et al. 2014, p. 375) - there has not been enough work to establish whether this is the case. The second issue with this debate is that explanations for why history should not matter in reputation building are premised on variables that only really apply to states, not non-state actors. For example, Tang (2005) argues that the anarchic features of the international system prevent reputations from forming. Additionally, imperfect and asymmetric informational environments mean that observers and audience members could form different reputational perspectives on actors or, alternatively, the audience members could take cues from each other on how their peers view an actor's reputation. The bearers of reputation belong to groups and this can influence their behavior, but observers of reputation also belong to groups that can influence their judgments.

In light of this tension in the IR literature, the organizational and labor economics literature has some profound and interesting insights on this.[18] Despite some claims from IR scholars that "game theorists and economists... these scholars are uninterested in how reputations form. Instead, they assume the conditions that must obtain for a reputation to form and then use the concept (as an independent variable) to solve problems" (Mercer 2010, p. 3), I would argue that more headway has been made in this literature than has been achieved in IR, mostly because these scholars are willing to consider more actors and

[18]The management and economics literature formalized the idea that reputations were assets firms developed to extract future rents (R. Wilson 1985) - to be "built, maintained, or "milked", and that crucially, reputation helped explain unexpected outcomes where reputation creates "inter-temporal linkages" between independent situations that ultimately "produce strategic and market behavior quite different from that predicted from analyses of the situations in isolation". The definition of reputation was accepted as history-dependent, where "the player's reputation is the history of his previously observed actions".

different types of reputations.

Tirole (1996) has been responsible for the most compelling and groundbreaking insights into how reputations form in his work on *A Theory of Collective Reputations*. Here, Tirole advances a formal model of the interdependence between collective and individual reputations, but also that reputation must necessarily be history-dependent because generations of actors overlap. In this model, reputation is history-dependent, however the referent's behavior could encompass a wide array of potential behaviors, not just an ability to maintain commitments. Tirole does outline the conditions under which reputation can be reset; i.e. amnesty, or revolution - where there is a complete overhaul and therefore no overlap in actors with reputations for corruption or cheating. One lasting implication of Tirole's model on generational reputations that seems to translate well to the literature on state reputations and IR is that reputations are sticky, and it is difficult to shake being tarred as a dishonest type (Mearsheimer 2011; Sartori 2002).

According to Tirole's model, agent's with reputations for corruption (or bad types) are still delegated tasks by principles since it is always optimal to delegate rather than abstain. However, once the agent has been caught cheating, the task given to the agent is down-graded from the high-impact task to the low impact task (task 1 to task 2) to minimize losses from corruption. One issue with this model for reputation in global politics is that Tirole does not consider whether some situations are higher or lower salience, such that the actor has the option of resetting their reputation. If the agent cheats or defects in a low-saliency situation, then the costs are minimized, while the agent performing in a high-salient situation may help restore the principal's faith. Tirole also does

not consider the relative reputation of the cheating agent vis-a-vi other available agents the principle may use. In Tirole's model, there is one agent and one principle. Thus, bearer's of reputation do not have peers. This type of model is difficult to apply to a vision of the world with multiple competing agents. This suits Tirole's formal model, and universes where principles must always engage the same pool of agents, however, what happens when we have many actors (states, firms, organizations) and many audiences?

History-dependence is less relevant in these situations, because audiences use heuristic short-cuts with the most accessible information to form ideas on an actor's reputation. For instance, evaluating firms by their nearest-neighbors, or industry-type. I bridge the gap between these two literatures by focusing on non-state actors (firms), and considering how a larger universe of competing agents, and a multiplicity of audiences changes the logic of horizontal (historical) versus vertical (nearest neighbor) assessments of reputation. Thus, rather than an either-or argument over history versus group comparisons, it might be the case that conditions over number of actors in a situation, salience of past events, the nature of the audience(s) and their ability to perceive signals, and the time-horizon of the actors involved, all factor into the extent to which history matters for reputation formation. While my argument here is that, for firms, nearest-neighbor behavior in a situation and observer comparisons within types matters more for formation of reputational types, this is not incompatible with the idea that for states (and international organizations), or situations involving one or a small pool of actors, reputations are history dependent.

With these considerations in mind, I offer a competing definition of reputation; *Reputation is beliefs that a group shares about an actor's behavior in a*

*setting*. This definition allows for the context and domain of behavior to shape which type of reputation matters, and while shared beliefs among the observers of a reputation is a necessary criteria of reputation, these beliefs might differ between groups.

# 6 A Group-Based Theory of Reputation

The starting point for my theory of firm reputation is that firm's necessarily possess multiple reputations because they have heterogeneous audiences (or stakeholders). There is always a hierarchy of reputational types, and this hierarchy determines which reputation a firm should invest in and when it makes sense to signal their type. When conditions are stable, firms prioritize catering to the audience they have transactions with that are essential to their survival over audiences that are affected by the company but not essential to their existence. For firms, we can imagine there are multiple stakeholders including customers, investors, suppliers and employees etc. At any point in time, firms must therefore manage conflicting expectations from audiences and maintain favorable reputation among all groups otherwise the company will lose its competitive advantage. The assumption is that this is easier when conditions are stable and its obvious to the firm which reputational type matters - and what signals matter.[19] However, when an exogenous shock occurs that causes the audience to pay attention to a different type of reputation that the firm has either not invested in, or the audience has unknown information about because previously they were not paying attention, the hierarchy of reputational types becomes more dynamic. In the short-term, it may be better for the firm to signal this newly salient reputational type. This has two implications; first,

[19]While the outline here suggests there are stable reputations over time and then shocks which create newly salient reputations, it's also the case that some reputational types may slowly become more important (more salient) to an audience over time, and this shift is distinctive but not immediate.

signaling this newly salient reputation among one audience may bring the firm into conflict with another audience. Previously, it may have been easier for the firm to have obscured their behavior, engaged in cheap-talk, or relied on the fact that the audience did not prioritize this reputational type. Thus firms face a trade-off between incentives to project this short-term type of reputation while potentially risking an unfavorable reputation among another audience. The second implication is the question of how observers assess firm type.

Both disciplinary literatures discussed here would suggest that under conditions of incomplete information about an actor's type, players gather information to form belief about another actor's type based on the actor's history. While slight deviations as to what history matters exists in both disciplines, reputation is thought to always be history-dependent. This assumption of history-dependence makes sense for highly salient reputations with key focal points (i.e. resolve as the reputation and conflict signaling/decisions to go to war as the focal point) where recall of the history of the actor is easy to access for the observer, making designation of their type almost intuitive. However, if this reputational type is relatively unknown among observers who may inconsistently or sporadically engage with the actor (i.e. consumers with certain firms), or there are too many focal points that are difficult to recall, or the situation itself is too distinct from past focal points, then audiences cannot rely on history to determine type. It is here that I argue for a group-based model of reputation.

What is meant by a group or collective in this context? Groups are not necessarily fixed, and their formation might depend on the context or the situation, however they are formed by alike actors who share a common set of distinctive traits. It's important that while the actor's traits sort them, it's the audience

themselves that are responsible for observing the group's existence. Within each group, these actors are subject to the same exogenous shock that occurs. In this context, the actor's within a group might have different types, but these are unknown to the observer. The collective group instead has a reputational benchmark - the pre-existing status quo of the group has now set the benchmark of appropriate behavior that the audience will accept. When audiences begin to pay attention to firm behavior associated with this now salient reputation, audiences will look to the behavior of the members of the collective to differentiate between types. If all actors remain within the same range of behavior, the status quo is the benchmark of reputational type and all actors are alike (that is they are both good and bad types). However, when one actor within the group departs from the status quo, it resets the benchmark of how audiences assess types. The early-moving actor has proven that departing from the status quo is possible, therefore they reveal both their type and other actors' types. This then forces the actor's within the group to make choices to follow the early-mover or remain at the previous status quo, which will inform audiences of their reputational type. Importantly, this adjustment of how observers assess type is dynamic. If, following this new assessment of types after the early-mover reacts, another player decides to go even further than the first early-mover - then the benchmark of what makes a good or bad type in this context is once again reset.

In some respects this model resembles Tomz (2012) who considers how borrowers differentiate their type to lenders by signaling whether they are lemons, fair-weathers, or stalwarts by repaying under certain conditions. However, my model departs from Tomz in three important respects. First, Tomz considers groups to be reputationally-derived; the borrower's reputation is their group, the borrower moves between groups based on how investors updates their be-

liefs about their type. In my model of reputation, groups are determined pre-reputational type; your group is whatever formal or informal collective you belong to based on easy-available criteria and how audiences think of "alike pairs"/"birds of a feather".[20] Second, and most importantly, investors update expectations based on a borrower's type (i.e., whether its default was strategic or unavoidable) based on the borrower's past and present behavior. The updates happen within a bilateral framework - the country and its creditors - not necessarily in a comparative way across a peer group. Reputation serves as a commitment mechanism - governments repay debts primarily to preserve their future borrowing ability, not because they are concerned about how their peers affect their standing among their immediate audience. Third, and relatedly, the criteria for a lemon, fair-weather, and stalwart remains constant - investors know which firm is which based on their actions in isolation. For Tomz, if a borrower changes their behavior, investors update their information about the borrower and they become a different type – they do not change the expectations of what it means to be a certain type. In my model, actor's within a group reset the benchmark of reputational types; using Tomz's debt repayment example, my model would mean that when a borrower that existed existed within one of the reputational types of groupings - stalwarts, fair-weathers or lemons – changed their debt repayment behavior, their actions created expectations on members of that reputational type to also shift their behavior because expectations of what was a stalwart, fair-weather, or lemon changed. This would mean that the benchmark or criteria to be a stalwart changes to match the most

[20] So, if you're a clothing company, your group is retail companies. Similarly, if you're an oil refinement company, your group is energy (oil and gas) companies. Audiences have different benchmarks within these groups – but you can't lose your group – you will always belong to that group no matter what you do, unless you cease to exist or you change businesses or the key criteria that the grouping is derived suddenly changes. In some ways this would also mean you cease to exist though; if you're a café company and you expand into food products, the version of the company that existed within café conglomerates cease to exist, and a new company now exists.

consistent borrower, and all other members previously considered stalwarts are now fair-weathers. These are the key differences between a group-based model of reputation and Tomz (2012) model of Bayesian updating and reputational types.

## 6.1 Reputation as a Dynamic Group Effect

There are few scholars who consider the role of peer behavior in reputation formation in IR. Crescenzi (2018) comes closest to considering group-behavior in his work on state reputations and alliance formation. Crescenzi (2018) argues that states make decisions on whether to go to war, or form an alliance, based on the reputation of other states with each other.[21] Crescenzi (2018) argument is that states form inferences based on the interactions of states with other states within a system, and through this states learn information that they use to form judgments over an actors reputational type (for Crescenzi, this is credibility in keeping commitments, and resolve for peace or alliance formation). Temporal proximity is also important; the closer the event that was used to generate information on the reputational type of the agent, the greater weight applied by the observing state. Crescenzi uses an illustration of state interactions between North Korea, the United States, and Iraq. Accordingly, North Korea uses the history of US interactions with Iraq as information on reputation. The greater the similarities between states in terms of policy and power resources, the more the state is able to use another state as a proxy for its likely relationship with states in a system. The inter-relationships of states used to inform reputation is interesting in this context, although Crescenzi (2018) does not go further in his argument that state's use alike proxies to make informed decisions.

[21] See also: (Henry 2022) on alliance interdependence, although Henry differs here by convincingly arguing that this results from developing a reputation for 'observed reliability', not resolve or royalty.

Crescenzi (2018) work is focused on the strategic behavior between states, where state A observers the strategic interactions of state B and C, state B is similar to themselves in terms of policy and/or power and it functions as a relevant proxy for state A's relationship with state C. State A updates their expectations about state C. Crescenzi (2018) is ultimately interested in showing that state's can build reputations for cooperation/resolve that affect alliance formation and conflict. The point, similar to Tomz (2012), is that contextual information and situational traits (such as repaying during crisis) help inform observer's of an actor's type. While Crescenzi (2018) considers temporally-proximate information as more relevant, his model still leans on repeated interactions and historical precedent where actors update beliefs, and there is no contagion effect of reputation - a state's reputation still exists in isolation from its peers.

Once again, historically-derived reputations make sense for highly salient reputations, such as reputations for resolve or quality. However, for less salient reputational types, such as firm reputations for being green, or human rights abiding reputations, observers may possess incomplete information about a firms type. This is because they either paid less attention to firm signals for these reputational types in the past because this reputational type was not salient, or the signals themselves were weak because the actor felt little incentive to signal their type given a lack of audience demand. In this context, observers can use benchmarks of behavior to help weigh indices which in turn are used to form judgments on behavior; those benchmarks are decided by the within-group behavior.

Groups are not necessarily fixed, and their formation might depend on the

context or the situation, however they are formed by alike actors who have something in common. These actors may be facing a similar problem of which they must all make a decision; or an actor has positioned themselves within this group by joining it or establishing themselves in opposition to this group. In both scenarios, the pre-existing status quo of the group has now set the benchmark of appropriate behavior that the audience will use to gain information about member firm's types. When one actor within the group departs from the status quo, it resets the benchmark, forcing the actor's within the group to make choices which will inform audiences of their reputational type.

When early-movers within these sectors depart from the status quo and leave or issue disclosures, they reset the benchmark of acceptable behavior among the audience. Audiences now form different expectations over the range of possible actions open to the actors in the group, and therefore actors that do not adjust to the new status quo send a signal to the audience that these actors are a different type. Because there is no pre-existing criteria for what makes a good or bad type, or observers are unsure what is feasible or possible given the constraints of the situation, if one firm from the group makes a decision to deviate from the group's status quo, it signals information about the relative criteria for firm type and resets the benchmark. The firm that moved may have acquired a good reputational type at the expense of its peers, although this effect is temporary for two reasons. First, the firm's peers could make the identical move, in which case the early-mover advantage is short-lived and all firm's now have the same reputational type, or another firm may go further and send a stronger signal, again resetting the benchmark for reputational types. Second, if audience attention to this reputational type wanes over time, then observers are less likely to pay attention to signals. From this vantage point, reputation

is a positional good where the rewards or benefits of maintaining a particular type of reputation are accrued not from how well the actor has performed in absolute terms, but rather the comparison of how the actor's behavior compares to their peers.

# 7 Pathways of Reputation: Cascades, Externalities, and Early-Movers

The mechanism through which reputation operates has been narrowed significantly in IR through the focus on resolve as the major reputational type. For these scholars, reputation is a mechanism for conveying information regarding a state's propensity for resolve, whether this be defense of honor as in Thucydides classic narrative of the Peloponnesian War, or willingness to retaliate and punish (T. C. Schelling 1956; T. Schelling 1960). Reputation is outlined as a mechanism in itself to encourage deterrence or signal resolve (Crescenzi 2018). However, there are many advantages to considering reputation as a group-based phenomenon for both firms and observers, since it opens up the possibility that there are group-effects and channels through which reputation may have contagion effects on peer behavior, and how and which audiences reward early-movers, and punish late-adopters. In addition, if observers themselves exist in groups and firms must appease all relevant audiences, sudden changes in signaling to one audience may impact the firm's relationship with another audience. Below I consider three pathways through which reputation can affect firm behavior in international politics: reputational cascades, early-mover incentives, and dueling audiences.

## 7.1 Cascades

Reputations are not formed in vacuums, and they rely on a tenuous informational environment. When a state's reputation is formed, augmented, destroyed or reconfigured in world politics, it is through a deliberative process among groups of individuals and states. I will examine how these informational and deliberative influences can produce 'reputational cascades'. Building on the definition offered by C. R. Sunstein (2019)'s work on social cascades, I define reputational cascades as large-scale judgments and statecraft movements in which actors "end up thinking something, or doing something, because of the beliefs or actions of a few early movers" (C. R. Sunstein 2002, p. 19). The concept of 'early movers' is not mainstream in IR discussions, despite how important it is for shaping policy. Take for example COVID-19; early-movers like China, South-Korea and Japan responded to the COVID-19 crisis with what would become known as 'lockdowns'. China's Hubei province locked-down in January 2020; an action that was swiftly followed by Italy, and a few days later Spain and France. The US's initial warning on travel to China on January 27th 2020 triggered Australia, Singapore, and Russia to ban travelers from China. The initial nature of responses to an unprecedented global pandemic was to watch how the unknown virus played out in first-hit countries, to observe how state's managed transnational and then national movements, and whether and how to implement restrictions. These cascades explain how legal, political and social movements can be rapid and produce tenuous situations.

One of the first studies finding evidence of social cascades was conducted by Salganik et al. (2006). In their study, participants would hear and download one or more of 72 songs and individuals were not told what anybody else had downloaded or their preferences. Eight other subgroups were created where in-

dividuals could view how many times each song was downloaded. The findings showed that the songs ranked best or worst by the control group never ended up at the top or bottom of the other subgroups, instead if a song received an early burst of downloads, it excelled in the rankings. Without the early movers pushing the song, any other song could fail (Salganik and Watts 2008; Salganik and Watts 2009). The effect of cascades extends to interpretations and judgments over facts as well. The judicial system of basing judgments on precedents functions as an informal institution of cascade effects, where decisions are based on the verdict of others rather than an individual judgments (Daughety and Reinganum 1999). An important study by Miceli and Coşgel (1994) shows how a judge's concerns for reputation informs, restrains and inspires judgments. In the authors' model, the preferences of judges were represented by a utility function with two components: private utility and reputational utility. The judge's private utility arose from his personal view of how a case should be decided, while his reputational utility depended on his expectation of how his decisions would be viewed by observers of the legal process: other judges, lawyers, litigants and the public.

Cascades can be divided into two types: informational cascades and reputational cascades. Informational cascades occur when "people silence themselves out of deference to the information conveyed by others" in contrast in a reputational cascade, individuals "silence themselves so as to avoid the opprobrium of others" (C. R. Sunstein 2006, p. 200). Silencing does not just mean the individual does not act/fails to speak up, but the individual goes against what they think to be right or what is likely to be right (i.e. silencing of internal preferences).[22]

[22]Social cascades are the umbrella term for both reputational and informational cascades.

A famous laboratory experiment on urns illustrates the effects of informational cascades (Anderson and Holt 1997). In this experiment, subjects had to guess whether the experiment was using an urn filled with two red balls and one white ball (urn A) or one containing two white balls and one red (urn B). For each period one of the urn's balls were emptied into a container, a randomly selected participant selected one ball, recording the ball's colour and their opinion about which urn was in use. The subject's colour was undisclosed to the group, but the subject's opinion on which urn was in use was shared. All participants would complete the same task and after the last participant, the answer of which urn was in use would be announced. The experiment revealed that individuals would make decisions about which urn was in use based on the majority of previous announcements, even if it contradicted with the color of the ball (Hung and Plott 2001). Over 77 percent of each round of the experiment revealed cascades, with few 'private signals', that is information provided by the individual's own draw. Interestingly, when one person's ball draw contradicted the decision of two or more previous participants, the next participant was likely to ignore this 'outlier' and follow the consensus of those who went before.[23] Although the majority of decisions followed Bayes law, the conclusion from the study was, "Initial misrepresentative signals start a chain of incorrect decisions that is not broken by more representative signals received later" (Anderson and Holt 1997, p. 847). Hung and Plott (2001) replicated a version of this experiment using rewards where subjects were paid \$2 for a correct group decision, and fined \$2 for an incorrect group decision, with the group decision based on the majority and no individual reward for individual decisions. In this experiment, just 39 percent of rounds witnessed cascades, and in 92 percent of cases, individual announcements matched their private ball draw. The group reward encouraged

[23] 72 percent of subjects followed Bayes's rule in the Anderson and Holt (1997) experiment, and 64 percent in Willinger and Ziegelmeyer (1998).

individuals to reveal private information. In the first experiment, participants did not care about sending an accurate signal to others, it was essentially an "informational externality" which did not affect an individual's gain, if the signal misled others, the participant had no real reason to care.

Reputational cascades are a distinct phenomenon to informational cascades, but we can think of them as siblings (Kuran 1998).[24] In an informational cascade individuals fail to reveal private information to the group out of concern that their information is 'wrong' because their information contradicts that of the first early-movers. In contrast, in reputational cascades the failure to reveal private information to the group does not arise out of a fear that their information is 'wrong' but rather the fear of exclusion or disapproval for contradicting the group. In order to preserve or project a particular reputation, individuals will speak out - or remain silent (C. R. Sunstein 1999, p. 8) by conforming to the first early-movers actions. Kuran (1998) refers to this as 'pluralistic ignorance'.[25]

Distinguishing between reputational and informational cascades is theoretically important. First, it explains why, even in the absence of private informa-

[24] Although conceptually similar, reputational cascades are distinct from that of the 'domino effect'. The domino effect refers to, "changes in one country's political institutions spread to neighboring countries, affecting these countries' political institutions similarly, which spreads to their neighbors, and so on. According to the democratic domino theory, for instance, increases in one nation's democracy lead to increases in its neighbors' democracy, leading to increases in their neighbors' democracy, and so on" (Leeson and A. M. Dean 2009, p. 534). Reputational cascades are distinctive in several ways. The domino effect refers to observed outcomes but has no mechanism - instead there are several speculated mechanisms that explain why, for example, neighboring countries would 'fall' to communism in succession (Simmons et al. 2006; Pevehouse 2002a; Pevehouse 2002b). In contrast a reputational cascade is a mechanism by definition. Second, the cascade effect is driven by concern for reputation, while the literature on the domino effect has never included reputation as a possible mechanism. Third, the logic of both are different: in a domino effect, the institutional structure of the actor changes not because of concern for observers but because of internal preferences, in a reputational cascade, actors are making decisions based on a) peer behavior and b) concern for how observer's view their reputation.

[25] For more on pluralistic ignorance see: J. Shamir and M. Shamir (1997), Miller and Prentice (1994), and J. O'Gorman (1986).

tion, actors will make decisions against their preferences or what they think to be the correct decision, out of fear of social sanctions. Second, informational cascades assume rational Bayesian updating - actors believe the early-mover(s) have the correct answer based on private information, so they update their beliefs of what is right. In a reputational cascade, actors may irrationally move contrary to their beliefs and private information, not because they think the early-mover has information on the optimal choice, but because they want to signal alignment and avoid social penalties. Unlike the informational cascade, the signal individuals send is intentional, designed to preserve status rather than to ensure informational accuracy in the group.

An example of a reputational cascade is given by C. Sunstein (2019, p. 68):

> "Suppose that Albert suggests that genetically modified foods are a serious problem and that Barbara concurs publicly with Albert, not because she actually agrees with Albert but because she does not wish to seem, to Albert, to be ignorant or indifferent to human health and environmental protection. If Albert and Barbara agree that genetically modified foods are a serious problem, Cynthia might not contradict them publicly and might even seem to share their judgment, not because she believes the judgment to be correct but because she does not want to face the hostility or lose the good opinion of others. It is easy to see how this process might generate a cascade. Once Albert, Barbara, and Cynthia offer a united front on the issue, their friend David might be most reluctant to contradict them even if he thinks they are wrong."

We can see how this would play out in IR. Suppose a state suggests that the actions of state B in dumping toxic nuclear waste is a serious problem, and that state C concurs publicly with state $A$, not because state $C$ necessarily agrees with state $A$ but because they do not want to appear to state $A$ as ignorant or indifferent to health or environmental protection. This effect is even greater for international organizations - for instance, thinking through the above scenario but replacing state $A$ with an IO. The cascade becomes clear if state $D$ and $E$ become more reluctant to contradict the united front of state $A$ and $B$, even if state $D$ or $E$ believe privately (and individually) that they are wrong.

Of course in world politics with deliberative decision-making by states and actors, other states are uncertain whether public statements are the product of independent knowledge, informational cascades or pressure induced by reputational concerns.[26] Legislators suffer from reputational cascades when they're pressured by party-members or the public, for instance consider the Conservative parties decision to hold a referendum on Brexit. Although we know there were few rebels who defected from the party, many more publicly supported Brexit despite private misgivings in order to avoid injuring their reputations.[27]

A reputational cascade is not a normative concept in and of itself, however they can lead to "widespread errors, factual or otherwise. Cascades need not involve deliberation; but related problems infect processes of group deliberation, as we will now see" (C. R. Sunstein 1999, p. 9). Rather convergence based on early-movers can lead actors to form distorted perceptions of another actor's reputational type, leading to inefficient outcomes (at best) and escalatory high-

[26] For corporate scandals, see: Thompson (2004).

[27] David Cameron himself admits he regrets how he handled the Brexit campaign, influenced by his concerns for his reputation within the party (with a vocal right-wing faction) and the public (with growing anti-European sentiments). See Also: Stimmer and Gliserman (2022).

risk situations (at worst). In some situations of IR a cascade effect will overcome collective torpor or reluctance in creating public goods, for instance by inducing firms to act on climate change. In other situations cascade effects will produce large-scale distortions in state-elite judgments, creating unjust and escalatory foreign policy responses as state's quickly align their public beliefs and positions with 'first-movers'. Another reason why the phenomenon of 'Reputational Cascades' occurs in IR is the high prevalence of risk (Kuran and C. R. Sunstein 1998).

Reputational cascades as a mechanism of reputation have largely been absent from the literature in IR, despite the profound applicability. Exceptions include very recent work by Hale (2022) to explain the rally around the flag affect around Putin, and Pearlman (2018) protests in Syria. Outside of political science, and the work of Sunstein himself, almost no work has considered companies as having network effects of reputation and the effect of reputational cascades on behavior. However, given the larger universe of firms than states, reputational cascades can be an interesting means through which concern for reputation influences firm behavior over issues like disclosing information, changing policies, moving operations, or even distancing the firm from other players. Below, I consider how this mechanism is triggered, through the role of early-movers.

## 7.2 Early-Movers

There are numerous reasons why the 'early-mover' may create reputational cascades. Firstly, one consequence of conclusion bias - that is, individuals have a tendency to collect, select or distort evidence to support a judgments they

already believe or desire to be true - is the 'anchoring effect'. The anchoring effect is "the disproportionate influence on decision makers to make judgments that are biased toward an initially presented value" (Furnham and Boo 2011, p. 35).[28] In their study, Tversky and Kahneman (1974) had participants estimate the percentage of African nations within the United Nations. The participants would spin a wheel with numbers between 0-100, and were then asked if their estimation of the quantity was higher or lower than the number, and then to provide an estimate by moving up or down from that number. Each group was given a different 'starting-point' based on the randomized spinning of the wheel. These arbitrary numbers had a significant effect on estimates (Tversky and Kahneman 1974, p. 1128). Anchoring is one of the most robust effects in cognitive heuristics. Studies have found the anchoring effect beyond quantitative numbers, and in the realm of general knowledge (Epley and Gilovich 2001; McElroy and Dowd 2007; Mussweiler and Englich 2005; Mussweiler and Strack 1999; Mussweiler 2001; Strack and Mussweiler 1997), probability estimates (Chapman and Johnson 1999; Plous 1989), legal judgments (Englich et al. 2005; Enough and Mussweiler 2001; Englich and Soder 2009; Strack and Mussweiler 1997), valuations and decision to purchase (Ariely et al. 2003; Wansink et al. 1998; Mussweiler et al. 2000), negotiations (Galinsky and Mussweiler 2001), and even self-efficacy (Cervone and Peake 1986). An event-related brain potential (ERP) experiment found electrophysiological evidence for the adjustment of estimates up and down from a generated anchor (Qu et al. 2008). A most interesting aspect of the anchoring effect that makes it especially relevant to reputation in IR is that the anchoring effect is stronger when there is a higher degree of ambiguity, a lower familiarity with the situation or the subject, personal involvement, and a more trustworthy anchoring bid (Van Exel et al. 2006). In IR uncertainty characterizes the international system, and many of these conditions are met

[28] See also: Tversky and Kahneman (1974).

amongst both the public and elites, making it a conducive area of study for anchoring.

Furthermore, additional research suggests that anchors with informational relevance to the situation or subject can have a stronger anchoring effect (Furnham and Boo 2011). In the above case, the number generated by the wheel was arbitrary and entirely random; subjects knew the wheel number and their estimate were separate entities and unrelated. The anchoring effect was significant in many of these studies, but it has led some critics to suggest that the anchoring effect is vulnerable to the relevance of the reference bid. In the case of IR, when 'first movers' come forward and generate an initial response to a situation, the resulting response becomes the 'anchor'. This type of anchor is directly related to the situation, even if the observer is biased, misinterprets the situation, or is being strategically misleading in issuing a response. As a result, we would expect the anchoring effect to be more effective in generating a 'first-mover' response as the reference bid is directly related to the estimates.[29]

Early-movers with extreme responses might lead to larger anchoring effects (Mussweiler and Strack 1999; Strack and Mussweiler 1997), although this is disputed by studies claiming implausibly extreme anchors lead to smaller anchoring effects (Wegener et al. 2001; Mussweiler and Strack 2001). Critics might suggest that political leaders and elites face stronger motivations to be accurate in their estimates, which may reduce the magnitude of the anchoring effect. However, even when offered payoffs for accuracy, Tversky and Kahneman (1974) and T. D. Wilson et al. (1996) find that incentives do not ameliorate the effect

[29] Although there is some interesting research that suggests elites will respond differently to the public from anchoring. T. D. Wilson et al. (1996) found that knowledgeable people are less affected by anchors. It's also the case that experts see themselves as less susceptible to anchoring than is justified by the results (Tversky and Kahneman 1974).

of anchoring.

Alternative explanations for cascade effects are persuasive arguments, and that actor's are merely operating within their own self-interest outside the effects of group-induced reputational concerns. We can disprove self-interest if we can show that actors move in a rapid direction towards an outcome that is either not clearly aligned to their interests or clashes with their interests.

## 7.3 Dueling Audiences

Defining reputation as a dynamic concept that fluctuates in reaction to peer-behavior has two implications; reputational externalities for collective reputations, and reputation stealing externalities for individuals. However the task of signaling reputational type in reaction to within-group movement is complicated by divergent audience expectations. I have already discussed how unitary audiences have been decomposed by IR scholars such as J. D. Kertzer and Brutger (2016) in highlighting how individual traits of the public can affect how observers perceive signals and value reputational types. However, for firms, relevant audiences extend beyond the public at large. Take for instance V. Rindova and Martins (2012, p. 1033) position on reputation as "stakeholders' perceptions of an organization's ability to create value relative to competitors". Audiences for firms comprise of stakeholders, examples of which include customers (Shapiro 1983), strategic alliances and partners (Dollinger et al. 1997), managers, investors, regulators, suppliers, and employees (C. Fombrun and Shanley 1990). These audiences have different expectations of the firm and different levels of knowledge of the firm's type. There is necessarily a hierarchy of audiences based on stakeholder priority.

One way of thinking of this is Clarkson (1995) differentiation between primary and secondary audiences. Primary audiences "have transactions with the company that are necessary for company survival" while secondary audiences are "affected by the company, but are not engaged in transactions with it and are not essential to its existence" (Carter and Deephouse 1999, p. 309). Consumers, investors and suppliers occupy the role of primary audiences of a firm. These groups have different expectations of the firm and as a result, multiple reputations can emerge. An example of this is given by Carter and Deephouse (1999), who finds conflicting reputations in firms sending price-cutting signals to build a predatory reputation towards competitors and new entrants,[30] and premium pricing signals to build a reputation for quality product or services among consumers.

An example of conflicting reputational types among observers is J. D. Kertzer and Brutger (2016) finding that Hawks reward reputations for resolve (or consistency) compared to Doves. However, in Kertzer & Brutger's example the actor creates their primary audience by signaling the threat of force where Doves punish reputations for belligerence. The actor cannot recover their losses among Doves by signaling follow-through, and therefore their survival (or just support) depends on meeting the expectations of the Hawks by signaling a reputation for consistency/follow-through. In the case of firms, observer groups are predefined, and while surely differences in interpretation of signals and value of reputational types among specific groups (i.e. consumers) exists, preferences around certain reputations is structured by the engagement of the observer with the firm.[31]

[30] See also Milgrom and J. Roberts (1982).

[31] Another feature of firms that make them distinctive to states is that the asset accumulation of firm reputation is determined by the size of their audience and relative attention they receive from their audience. In this way, larger firms with more established brands have larger audiences, are reputational costs and gains are larger (V. Rindova and Martins 2012).

In the absence of high-salience issues, managing multiple reputations is easier as the type of signal the firm sends remains stable. However, when an exogenous shock increases the attention of audiences towards corporate actors who are either directly or indirectly implicated - this could be an economic, social, environmental, or geopolitical crisis - reputational preferences are destabilized and firm behavior in a specific area becomes a focal point. While firms may have previously felt little pressure to project a socially-conscious, or environmentally-conscious reputational type, during periods of heightened awareness, companies may respond to consumer or activist attention by leveraging this opportunity to build reputational capital (Gregorich et al. 2023). The trade-off, however, is that signaling this reputational type can alienate other groups of stakeholders that may oppose the firm's behavior, especially if it threatens to undermine what they value in the firm (Burbano 2021), and they may react by imposing reputational costs.

It is this element of *observer group dynamics* in a theory of group-based reputation that creates a tension for whether firms respond to the crisis by signaling their type, or maintaining the status quo. As I suggest above, if an early-mover in the firm's cohort adjusts their behavior, they may be forced to follow suit. An observable implication of this would be increased attention to appeasing and managing reputations among stakeholders who may have conflicting responses to this new reputational type. While this is speculative, one way that firm's can manage conflicting reputations among audience groups is by exploiting the informational heterogeneity across observer groups. Consumers may be less likely to process particular signals, consider only some informational cues, while investors may be more motivated to stay attuned to specific details

and evaluations of the firm Wang (2022). The result is that it may be easier for firms to manage conflicting reputations because it's easier for them to engage in 'cheap talk' with consumers while sending costly signals to investors. I leave this line of inquiry to future research.

# 8 Conclusion

Corporations are increasingly adopting governance tasks and powers, from control over transportation, the flow of information, development and control over new forms of energy, and issuing new currencies. The size of many corporations has outpaced that of a modern state, and they participate in politics more effectively than ordinary citizens. Though economists continue to rally against the politicization of the firm, and implored corporations to "remind stakeholders that the business of the business is, well, business" (Casey and Ginsburg 2025), that pursuit however is complicated when firms are either the cause or solution to political problems. From emissions to the imposition of economic sanctions, governments may require firm cooperation over governance problems. Unlike voters, however, firms are not elected into positions of power by the public, and government's capacity to regulate these actors is complicated by the complexity of multi-nationals, and the lobbying capacity of the firm. As I have argued in this paper, one mechanism through which firm's can be forced to self-regulate or change their own behavior is reputation.

This paper re-examines the role of reputation in influencing firm behavior in politics from the lens of two disciplines; micro-economics and management which has traditionally focused on firms in markets, and international relations, which has focused on the role of states under conditions of anarchy. I show that the structural conditions of the market and the firm in economics, and

the lingering focus on states and resolve in IR, led to a divergence in how each discipline dealt with the notion of reputation and reputational costs. I argue that the history-dependence assumption underpins both disciplines, and makes the assumption that all observers have i) no informational asymmetries, where observers know the history of the agent or that this history was transparent and easy to recall, and ii) that the history of the agent is cumulative, and that every observer places the same weight on each action of the agent and therefore the agent's behavior at a single time-point is somehow less informative than their behavior at past time-points. The implication is also that observers, not bearers, of reputation make judgments about their type and signaling independently from each other.

While the assumption of history-dependence likely holds for highly salient reputational types that are critical to an actor's survival - quality for firms and resolve for states - where observers engage with the actor's signals of this type in either more frequency or attention, the assumption is less stable when the reputational type is less salient, and informational-asymmetries of observers is high. Under these conditions, I argue that a group-based model of reputation where reputations are formed in relation to peers – reputations are bench-marked to an equilibrium of behavior established through a formal or informal group. Audiences group firms by some type of defining characteristic, and use this to gauge the behavior of the firm. When one actor moves in reaction to an exogenous event, they reset the boundaries of what is a 'good' or 'bad' type in whatever reputational domain the action is relevant to. Actors within that group update their beliefs about their own reputational type and the criteria to be a 'good' or 'bad' type. They then move to appease the audience.

A group-based model of reputation brings back the social dimension to reputation - neither actors or observers exist in a vacuum and reputational types develop in response to demand. A group-based model of reputation has important implications for the temporal-dimension of reputation and how it distributes across time and observer attention wanes, as well as how actor's deal with conflicting reputational types, and the pathways through which reputational effects diffuse - including cascades and early-mover effects.

These theoretical insights have implications for understanding how consumers can influence firm behavior and when self-regulation may be possible. They also suggest future lines of inquiry for research on the role of transparency in moderating the effects of reputation, as well as a demand-side examination of when consumers mobilize effectively to impose reputational costs - or whether consumer rhetoric versus behavior is misaligned. Just as IR moved to breakdown the black-box of the state and assign reputational stickiness to leaders, independent of the state, future research could extend this line of thinking to firm executives as leader-types in international politics. Exploring how firm executive's build reputations, and how executive's concern for reputation could influence how they build reputations vis-a-vi state leaders, regulators, or the public.

# References


Aaker, D. A. (2012). *Building strong brands*. Simon and schuster.

Albert, S. and D. A. Whetten (1985). "Organizational identity." In: *Research in organizational behavior*.

Anderson, L. R. and C. A. Holt (1997). "Information cascades in the laboratory". In: *The American economic review*, pp. 847–862.

Ariely, D., G. Loewenstein, and D. Prelec (2003). ""Coherent arbitrariness": Stable demand curves without stable preferences". In: *The Quarterly journal of economics* 118.1, pp. 73–106.

Avant, D. (2004). "The privatization of security and change in the control of force". In: *International studies perspectives* 5.2, pp. 153–157.

Backer, L. C. (2015). "Regulating multinational corporations: Trends, challenges, and opportunities". In: *Brown J. World Aff.* 22, p. 153.

Bell, M. S. and K. Quek (2018). "Authoritarian public opinion and the democratic peace". In: *International Organization* 72.1, pp. 227–242.

Berle, A. and G. C. Means (1933). *The Modern Corporation and Private Property*. Vol. 1. Macmillan.

Bradly, A. and G. Nathan (2019). "Institutional CSR: provision of public goods in developing economies". In: *Social Responsibility Journal* 15.7, pp. 874–887.

Brown, J. R. and J. Huang (2020). "All the president's friends: Political access and firm value". In: *Journal of Financial Economics* 138.2, pp. 415–431.

Brutger, R. and J. D. Kertzer (2018). "A dispositional theory of reputation costs". In: *International Organization* 72.3, pp. 693–724.

Burbano, V. C. (2021). "The demotivating effects of communicating a social-political stance: Field experimental evidence from an online labor market platform". In: *Management Science* 67.2, pp. 1004–1025.

Carter, S. M. and D. L. Deephouse (1999). "'Tough talk'and 'soothing speech': Managing reputations for being tough and for being good". In: *Corporate Reputation Review* 2, pp. 308–332.

Casey, A. J. and T. Ginsburg (Jan. 2025). "Corporate Leaders Need to Keep Their Mouths Shut". In: *The New York Times*. Opinion. URL: https://www.nytimes.com/2025/01/29/opinion/speech-universities-corporations-executives.html.

Casler, D., D. Ribar, and K. Yarhi-Milo (2023). "The Many Faces of Credibility: Hawks, Doves, and Nuclear Disarmament". In: *Security Studies* 32.3, pp. 413–445.

Caves, R. E. and M. E. Porter (1977). "From entry barriers to mobility barriers: Conjectural decisions and contrived deterrence to new competition". In: *The quarterly journal of economics* 91.2, pp. 241–261.

Cervone, D. and P. K. Peake (1986). "Anchoring, efficacy, and action: The influence of judgmental heuristics on self-efficacy judgments and behavior." In: *Journal of Personality and social Psychology* 50.3, p. 492.

Chapman, G. B. and E. J. Johnson (1999). "Anchoring, activation, and the construction of values". In: *Organizational behavior and human decision processes* 79.2, pp. 115–153.

Clare, J. and V. Danilovic (2010). "Multiple audiences and reputation building in international conflicts". In: *Journal of Conflict Resolution* 54.6, pp. 860–882.

Clarkson, M. E. (1995). "A stakeholder framework for analyzing and evaluating corporate social performance". In: *Academy of management review* 20.1, pp. 92–117.

Connelly, B. L. et al. (2011). "Signaling theory: A review and assessment". In: *Journal of management* 37.1, pp. 39–67.

Crescenzi, M. J. (2018). *Of friends and foes: reputation and learning in international politics*. Oxford University Press.

Dafoe, A., J. Renshon, and P. Huth (2014). "Reputation and status as motives for war". In: *Annual Review of Political Science* 17, pp. 371–393.

Danilovic, V. (2002). *When the stakes are high: Deterrence and conflict among major powers*. University of Michigan Press.

Danilovic, V., J. Clare, and C. Tucker (2020). "Reputation in Deterrence and Decision-Making". In: *Oxford Research Encyclopedia of International Studies*.

Daughety, A. F. and J. F. Reinganum (1999). "Stampede to judgement: persuasive influence and herding behavior by courts". In: *American Law and Economics Review* 1.1, pp. 158–189.

Davies, G. A. and R. Johns (2013). "Audience costs among the British public: the impact of escalation, crisis type, and prime ministerial rhetoric". In: *International Studies Quarterly* 57.4, pp. 725–737.

Dean, M. (2021). "An Environmental FOIA". In: *California Law Review* 109.6, pp. 2423–2463.

Diehl, P. F. (1992). "What are they fighting for? The importance of issues in international conflict research". In: *Journal of Peace Research* 29.3, pp. 333–344.

Dollinger, M. J., P. A. Golden, and T. Saxton (1997). "The effect of reputation on the decision to joint venture". In: *Strategic management journal* 18.2, pp. 127–140.

Dowling, G. (2000). *Creating corporate reputations: Identity, image and performance: Identity, image and performance*. OUP Oxford.

Dowling, G. R. (2004). "Corporate reputations: should you compete on yours?" In: *California management review* 46.3, pp. 19–36.

Durand, R. and R. F. Haans (2022). "Optimally distinct? Understanding the motivation and ability of organizations to pursue optimal distinctiveness (or not)". In: *Organization Theory* 3.1, p. 26317877221079341.

Dutton, J. E. and J. M. Dukerich (1991). "Keeping an eye on the mirror: Image and identity in organizational adaptation". In: *Academy of management journal* 34.3, pp. 517–554.

Englich, B., T. Mussweiler, and F. Strack (2005). "The last word in court—A hidden disadvantage for the defense". In: *Law and Human Behavior* 29, pp. 705–722.

Englich, B. and K. Soder (2009). "Moody experts—How mood and expertise influence judgmental anchoring". In: *Judgment and Decision making* 4.1, pp. 41–50.

Enough, B. and T. Mussweiler (2001). "Sentencing under uncertainty: Anchoring effects in the courtroom 1". In: *Journal of applied social psychology* 31.7, pp. 1535–1551.

Epley, N. and T. Gilovich (2001). "Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors". In: *Psychological science* 12.5, pp. 391–396.

Falchetti, D., G. Cattani, and S. Ferriani (2022). "Start with "Why," but only if you have to: The strategic framing of novel ideas across different audiences". In: *Strategic Management Journal* 43.1, pp. 130–159.

Fearon, J. D. (1994). "Domestic political audiences and the escalation of international disputes". In: *American political science review* 88.3, pp. 577–592.

— (1995). "Rationalist explanations for war". In: *International organization* 49.3, pp. 379–414.

Fombrun, C. and M. Shanley (1990). "What's in a name? Reputation building and corporate strategy". In: *Academy of management Journal* 33.2, pp. 233–258.

Fombrun, C. J. (1996). "Reputation: Realizing value from the corporate image". In: *Harvard Business School*.

— (2005). "Corporate reputations as economic assets". In: *The Blackwell handbook of strategic management*, pp. 285–308.

Furnham, A. and H. C. Boo (2011). "A literature review of the anchoring effect". In: *The journal of socio-economics* 40.1, pp. 35–42.

Galinsky, A. D. and T. Mussweiler (2001). "First offers as anchors: the role of perspective-taking and negotiator focus." In: *Journal of personality and social psychology* 81.4, p. 657.

Gelpi, C. and J. M. Grieco (2001). "Attracting trouble: Democracy, leadership tenure, and the targeting of militarized challenges, 1918-1992". In: *Journal of Conflict Resolution* 45.6, pp. 794–817.

Gigerenzer, G. and W. Gaissmaier (2011). "Heuristic decision making". In: *Annual review of psychology* 62.2011, pp. 451–482.

Goldfien, M. A., M. F. Joseph, and R. W. McManus (2023). "The Domestic Sources of International Reputation". In: *American Political Science Review* 117.2, pp. 609–628.

Gregorich, G., V. Burbano, and D. Wang (2023). *Why firms speak up: Evidence from Black Lives Matter and Stop Asian Hate*. Tech. rep. Working Paper.

Guzman, A. T. (2002). "A compliance-based theory of international law". In: *Calif. L. Rev.* 90, p. 1823.

Hale, H. E. (2022). "Authoritarian rallying as reputational cascade? Evidence from Putin's popularity surge after Crimea". In: *American Political Science Review* 116.2, pp. 580–594.

Hall, R. (2009). "The strategic analysis of intangible resources". In: *Knowledge and strategy*. Routledge, pp. 181–195.

Heil, O. and T. S. Robertson (1991). "Toward a theory of competitive market signaling: A research agenda". In: *Strategic Management Journal* 12.6, pp. 403–418.

Henderson, D. (2001). *Misguided Virtue: False Notifications of Corporate Social Responsibility*.

Henry, I. D. (2022). *Reliability and Alliance Interdependence: The United States and Its Allies in Asia, 1949–1969*. Cornell University Press.

Hopf, T. (1994). *Peripheral visions: Deterrence theory and American foreign policy in the third world, 1965-1990*. University of Michigan Press.

Horowitz, M. C. and A. C. Stam (2014). "How prior military experience influences the future militarized behavior of leaders". In: *International Organization* 68.3, pp. 527–559.

Horwitt, D. (2022). "Fracking with "Forever Chemicals" in OHIO". In.

Hung, A. A. and C. R. Plott (2001). "Information cascades: Replication and an extension to majority rule and conformity-rewarding institutions". In: *American Economic Review* 91.5, pp. 1508–1520.

Huth, P. K. (1988). "Extended deterrence and the prevention of war". In: *(No Title)*.

— (1997). "Reputations and deterrence: A theoretical and empirical assessment". In: *Security Studies* 7.1, pp. 72–99.

J. O'Gorman, H. (1986). "The discovery of pluralistic ignorance: An ironic lesson". In: *Journal of the History of the Behavioral Sciences* 22.4, pp. 333–347.

Jensen, M. and A. Roy (2008). "Staging exchange partner choices: When do status and reputation matter?" In: *Academy of Management journal* 51.3, pp. 495–516.

Jensen, M. C. (2002). "Value maximization, stakeholder theory, and the corporate objective function". In: *Business ethics quarterly*, pp. 235–256.

Jervis, R. (1989). *The logic of images in international relations*. Columbia University Press.

— (2017a). *How statesmen think: The psychology of international politics*. Princeton University Press.

— (2017b). *Perception and misperception in international politics: New edition*. Princeton University Press.

Jervis, R., R. N. Lebow, and J. G. Stein (1985). *Psychology and deterrence*. JHU Press.

Jost, T. et al. (2024). "Advisers and Aggregation in Foreign Policy Decision Making". In: *International organization* 78.1, pp. 1–37.

Kahneman, D., P. Slovic, and A. Tversky (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge university press.

Kahneman, D. and A. Tversky (1972). "Subjective probability: A judgment of representativeness". In: *Cognitive psychology* 3.3, pp. 430–454.

Kertzer, J. (2016). *Resolve in international politics*. Vol. 2. Princeton University Press.

Kertzer, J. D. (2022). "Re-assessing elite-public gaps in political behavior". In: *American Journal of Political Science* 66.3, pp. 539–553.

Kertzer, J. D. and R. Brutger (2016). "Decomposing audience costs: Bringing the audience back into audience cost theory". In: *American Journal of Political Science* 60.1, pp. 234–249.

Kertzer, J. D., R. Brutger, and K. Quek (2019). *Perspective taking and the security dilemma: Cross-national experimental evidence from China and the United States*. Tech. rep. Technical report. Harvard University.

Kertzer, J. D., J. Renshon, and K. Yarhi-Milo (2021). "How Do Observers Assess Resolve?" In: *British Journal of Political Science* 51.1, pp. 308–330.

Kim, H. and M. Jensen (2014). "Audience heterogeneity and the effectiveness of market signals: How to overcome liabilities of foreignness in film exports?" In: *Academy of Management Journal* 57.5, pp. 1360–1384.

King, B. G. and D. A. Whetten (2008). "Rethinking the relationship between reputation and legitimacy: A social actor conceptualization". In: *Corporate Reputation Review* 11, pp. 192–207.

Kreps, D. M. and R. Wilson (1982). "Reputation and imperfect information". In: *Journal of economic theory* 27.2, pp. 253–279.

Kreps, D. M. et al. (1982). "Rational cooperation in the finitely repeated prisoners' dilemma". In: *Journal of Economic theory* 27.2, pp. 245–252.

Kuran, T. (1998). *Private truths, public lies: The social consequences of preference falsification*. Harvard University Press.

Kuran, T. and C. R. Sunstein (1998). "Availability cascades and risk regulation". In: *Stan. L. Rev.* 51, p. 683.

Lambert, T. (2019). "Lobbying on regulatory enforcement actions: Evidence from US commercial and savings banks". In: *Management Science* 65.6, pp. 2545–2572.

Leeson, P. T. and A. M. Dean (2009). "The democratic domino theory: An empirical investigation". In: *American Journal of Political Science* 53.3, pp. 533–551.

Levendusky, M. S. and M. C. Horowitz (2012). "When backing down is the right decision: Partisanship, new information, and audience costs". In: *The Journal of Politics* 74.2, pp. 323–338.

Madsen, P. M. and Z. J. Rodgers (2015). "Looking good by doing good: The antecedents and consequences of stakeholder attention to corporate disaster relief". In: *Strategic Management Journal* 36.5, pp. 776–794.

Manyika, J. and M. Birshan (2021). "Companies in the 21st century A new look at how corporations impact the economy and households". In.

Matten, D. (2009). "Review essay:"It's the Politics, Stupid!" Reflections on the role of business in contemporary nonfiction". In: *Business & Society* 48.4, pp. 565–576.

McElroy, T. and K. Dowd (2007). "Susceptibility to anchoring effects: How openness-to-experience influences responses to anchoring cues". In: *Judgment and Decision making* 2.1, pp. 48–53.

Mearsheimer, J. J. (2011). *Why leaders lie: The truth about lying in international politics*. Oxford University Press.

Meckling, W. H. and M. C. Jensen (1976). "Theory of the Firm". In: *Managerial behavior, agency costs and ownership structure* 3.4, pp. 305–360.

Mercer, J. (2010). *Reputation and international politics*. Cornell University Press.

Miceli, T. J. and M. M. Coşgel (1994). "Reputation and judicial decision-making". In: *Journal of Economic Behavior & Organization* 23.1, pp. 31–51.

Milgrom, P. and J. Roberts (1982). "Predation, reputation, and entry deterrence". In: *Journal of economic theory* 27.2, pp. 280–312.

Miller, D. T. and D. A. Prentice (1994). "Collective errors and errors about the collective". In: *Personality and Social Psychology Bulletin* 20.5, pp. 541–550.

Mussweiler, T. (2001). "The durability of anchoring effects". In: *European Journal of Social Psychology* 31.4, pp. 431–442.

Mussweiler, T. and B. Englich (2005). "Subliminal anchoring: Judgmental consequences and underlying mechanisms". In: *Organizational Behavior and Human Decision Processes* 98.2, pp. 133–143.

Mussweiler, T. and F. Strack (1999). "Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model". In: *Journal of Experimental Social Psychology* 35.2, pp. 136–164.

— (2001). "Considering the impossible: Explaining the effects of implausible anchors". In: *Social Cognition* 19.2, pp. 145–160.

Mussweiler, T., F. Strack, and T. Pfeiffer (2000). "Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility". In: *Personality and Social Psychology Bulletin* 26.9, pp. 1142–1150.

Myers, S. C. and N. S. Majluf (1984). "Corporate financing and investment decisions when firms have information that investors do not have". In: *Journal of financial economics* 13.2, pp. 187–221.

Negro, G., M. T. Hannan, and M. Fassiotto (2015). "Category signaling and reputation". In: *Organization Science* 26.2, pp. 584–600.

Pearlman, W. (2018). "Moral identity and protest cascades in Syria". In: *British Journal of Political Science* 48.4, pp. 877–901.

Pevehouse, J. C. (2002a). "Democracy from the outside-in? International organizations and democratization". In: *International organization* 56.3, pp. 515–549.

— (2002b). "With a little help from my friends? Regional organizations and the consolidation of democracy". In: *American Journal of Political Science*, pp. 611–626.

Plous, S. (1989). "Thinking the unthinkable: The effects of anchoring on likelihood estimates of nuclear war 1". In: *Journal of Applied Social Psychology* 19.1, pp. 67–91.

Podolny, J. M. (1993). "A status-based model of market competition". In: *American journal of sociology* 98.4, pp. 829–872.

— (1994). "Market uncertainty and the social character of economic exchange". In: *Administrative science quarterly*, pp. 458–483.

Pollock, T. G. and R. Gulati (2007). "Standing out from the crowd: The visibility-enhancing effects of IPO-related signals on alliance formation by entrepreneurial firms". In: *Strategic Organization* 5.4, pp. 339–372.

Press, D. G. (2005). *Calculating credibility: How leaders assess military threats*. Cornell University Press.

Qu, C., L. Zhou, and Y.-j. Luo (2008). "Electrophysiological correlates of adjustment process in anchoring effects". In: *Neuroscience Letters* 445.3, pp. 199–203.

Renshon, J., A. Dafoe, and P. Huth (2018). "Leader influence and reputation formation in world politics". In: *American Journal of Political Science* 62.2, pp. 325–339.

Renshon, J., K. Yarhi-Milo, and J. D. Kertzer (2023). "Democratic reputations in crises and war". In: *The Journal of Politics* 85.1, pp. 1–18.

Rescue, R. B. (2024). *Companies that are worth more than countries*.

Richard, L. (1981). *Between Peace and War: The Nature of International Crisis*.

Riley, T. (July 2017). "Just 100 Companies Responsible for 71% of Global Emissions, Study Says". In: *The Guardian*. Accessed: 2025-01-02. URL: https://www.theguardian.com/sustainable-business/2017/jul/10/100-fossil-fuel-companies-investors-responsible-71-global-emissions-cdp-study-climate-change.

Rindova, V. and L. L. Martins (2012). "Show Me the Money: A Multidimensional Perspective on Reputation as an Intangible Asset". In: *The Oxford Handbook of Corporate Reputation*. Ed. by M. L. Barnett and T. G. Pollock. Oxford, UK: Oxford University Press, pp. 16–33.

Rindova, V. P. and C. J. Fombrun (1999). "Constructing competitive advantage: the role of firm–constituent interactions". In: *Strategic management journal* 20.8, pp. 691–710.

Roberts, P. W. and G. R. Dowling (2002). "Corporate reputation and sustained superior financial performance". In: *Strategic management journal* 23.12, pp. 1077–1093.

Salganik, M. J., P. S. Dodds, and D. J. Watts (2006). "Experimental study of inequality and unpredictability in an artificial cultural market". In: *science* 311.5762, pp. 854–856.

Salganik, M. J. and D. J. Watts (2008). "Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market". In: *Social psychology quarterly* 71.4, pp. 338–355.

— (2009). "Web-based experiments for the study of collective social dynamics in cultural markets". In: *Topics in cognitive science* 1.3, pp. 439–468.

Sartori, A. E. (2002). "The might of the pen: A reputational theory of communication in international disputes". In: *International Organization* 56.1, pp. 121–149.

Schelling, T. (1960). "The Strategy of Conflict (Cambridge, MA: Harvard UP)". In.

Schelling, T. C. (1956). "An essay on bargaining". In: *The American Economic Review* 46.3, pp. 281–306.

— (1967). "Arms and Influence (3 leid.)" In: *New Haven, London: Yale*.

Shamir, J. and M. Shamir (1997). "Pluralistic ignorance across issues and over time: Information cues and biases". In: *Public Opinion Quarterly*, pp. 227–260.

Shapira, R. and L. Zingales (2017). *Is pollution value-maximizing? The DuPont case*. Tech. rep. National Bureau of Economic Research.

Shapiro, C. (1983). "Premiums for high quality products as returns to reputations". In: *The quarterly journal of economics* 98.4, pp. 659–679.

Sharkey, A., B. Kovács, and G. Hsu (2023). "Expert critics, rankings, and review aggregators: The changing nature of intermediation and the rise of markets with multiple intermediaries". In: *Academy of Management Annals* 17.1, pp. 1–36.

Sharman, J. C. (2007). "Rationalist and constructivist perspectives on reputation". In: *Political Studies* 55.1, pp. 20–37.

Simmons, B. A., F. Dobbin, and G. Garrett (2006). "Introduction: The international diffusion of liberalism". In: *International organization* 60.4, pp. 781–810.

Snyder, G. H. and P. Diesing (2015). *Conflict among nations: Bargaining, decision making, and system structure in international crises*. Princeton University Press.

Spence, M. (1974). "Competitive and optimal responses to signals: An analysis of efficiency and distribution". In: *Journal of Economic theory* 7.3, pp. 296–332.

— (1978). "Job market signaling". In: *Uncertainty in economics*. Elsevier, pp. 281–306.

Stigler, G. J. (1962). "Information in the labor market". In: *Journal of political economy* 70.5, Part 2, pp. 94–105.

Stimmer, A. and J. Gliserman (2022). "Disentangling norms, morality, and principles: the September 2019 Brexit rebellion". In: *European Journal of International Relations* 28.4, pp. 955–982.

Strack, F. and T. Mussweiler (1997). "Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility." In: *Journal of personality and social psychology* 73.3, p. 437.

Sundaram, A. K. and A. C. Inkpen (2004). "The corporate objective revisited". In: *Organization science* 15.3, pp. 350–363.

Sunstein, C. (2019). *Conformity: The Power of Social Influences*. New York University Press.

Sunstein, C. R. (1999). "The law of group polarization". In: *University of Chicago Law School, John M. Olin Law & Economics Working Paper* 91.

— (2002). "Conformity and dissent". In: *Available at SSRN 341880*.

— (2006). "Deliberating groups versus prediction markets (or Hayek's challenge to Habermas)". In: *Episteme* 3.3, pp. 192–213.

— (2019). *How change happens*. Mit Press.

Tang, S. (2005). "Reputation, cult of reputation, and international conflict". In: *Security Studies* 14.1, pp. 34–62.

Thompson, L. D. (2004). *Corporate Scandals, why they happened and why they may not happen again*. Chautauqua Institution.

Tingley, D. H. and B. F. Walter (2011). "The effect of repeated play on reputation building: an experimental approach". In: *International Organization* 65.2, pp. 343–365.

Tirole, J. (1996). "A theory of collective reputations (with applications to the persistence of corruption and to firm quality)". In: *The Review of Economic Studies* 63.1, pp. 1–22.

Tomz, M. (2007). "Domestic audience costs in international relations: An experimental approach". In: *International Organization* 61.4, pp. 821–840.

— (2012). *Reputation and international cooperation: sovereign debt across three centuries*. Princeton University Press.

Treisman, D. (2004). "Rational appeasement". In: *International Organization* 58.2, pp. 345–373.

Tversky, A. and D. Kahneman (1974). "Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty." In: *science* 185.4157, pp. 1124–1131.

Van Exel, N. et al. (2006). "With a little help from an anchor: discussion and evidence of anchoring effects in contingent valuation". In: *The journal of socio-economics* 35.5, pp. 836–853.

Walsh, G. and S. E. Beatty (2007). "Customer-based corporate reputation of a service firm: scale development and validation". In: *Journal of the academy of marketing science* 35, pp. 127–143.

Waltz, K. (1979). *Theory of International Politics*. Vol. 1. Waveland Press.

Wang, P. (2022). "Looking into the past: Audience heterogeneity and the inconsistency of market signals". In: *Strategic Organization*, p. 14761270221139760.

Wansink, B., R. J. Kent, and S. J. Hoch (1998). "An anchoring and adjustment model of purchase quantity decisions". In: *Journal of Marketing Research* 35.1, pp. 71–81.

Wartick, S. L. (2002). "Measuring corporate reputation: Definition and data". In: *Business & Society* 41.4, pp. 371–392.

Wegener, D. T. et al. (2001). "Implications of attitude change theories for numerical anchoring: Anchor plausibility and the limits of anchor effectiveness". In: *Journal of Experimental Social Psychology* 37.1, pp. 62–69.

Weigelt, K. and C. Camerer (1988). "Reputation and corporate strategy: A review of recent theory and applications". In: *Strategic management journal* 9.5, pp. 443–454.

Weisiger, A. and K. Yarhi-Milo (2015). "Revisiting reputation: How past actions matter in international politics". In: *International Organization* 69.2, pp. 473–495.

Willinger, M. and A. Ziegelmeyer (1998). "Are more informed agents able to shatter information cascades in the lab?" In: *The economics of networks: Interaction and behaviours*. Springer, pp. 291–305.

Wilson, R. (1985). "Reputations in games and markets". In: *Game-theoretic models of bargaining* 96, pp. 27–62.

Wilson, T. D. et al. (1996). "A new look at anchoring effects: basic anchoring and its antecedents." In: *Journal of Experimental Psychology: General* 125.4, p. 387.

Wolford, S. (2007). "The turnover trap: New leaders, reputation, and international conflict". In: *American Journal of Political Science* 51.4, pp. 772–788.

Yarhi-Milo, K. (2018). *Who fights for reputation: The psychology of leaders in international conflict*. Vol. 156. Princeton University Press.

Zingales, L. (2017). "Towards a political theory of the firm". In: *Journal of Economic Perspectives* 31.3, pp. 113–130.

# Who Gives a Damn About a Bad Reputation? The Corporate Exodus and the Russian-Ukrainian War

Hayley Pring

Paper Two

## Abstract

Following Russia's invasion of Ukraine in 2022, some multinational corporations (MNCs) went beyond compliance with sanctions, and voluntarily withdrew from Russia, while other firms paused operations or continued 'business as usual'. Why did some MNCs issue statements to the public on their decision to leave or stay? I make two key arguments. First, I argue that early movers create a benchmark against which consumers and other observers measured their industry peers. Thus, when a precedent for leaving is set in an industry, those who don't follow suit will face a higher reputational costs of remaining. I test the idea of 'intra-industry diffusion' using cox proportional hazard models with a novel longitudinal dataset of collected statements issued by firms on their position in Russia across multiple time-points. Second, remaining firms attempted to appease both international and Russian audiences by projecting politically-neutral reputations. These firms employed specific strategies to mitigate reputational costs. Using word-embedding techniques and topic modeling on a corpus of company statements, I show that remaining firms are less critical of Russia, less likely to issue social media statements, and more likely to leverage the legitimacy of membership to UN-adjacent multi-stakeholder initiatives. I find that remaining

firms are also more likely to attempt to join these public-private partnerships post-invasion, suggesting bad actors learn how to leverage legitimacy from IO-adjacent organizations. This study provides insights into how reputation affects corporate responses to geopolitical shocks.

## 1 Introduction

Woodrow Wilson once called sanctions "something more tremendous than war".[1] Their use dates back to at least the 1815 Congress of Vienna, when British Foreign Secretary Castelreagh suggested boycotting colonial products to enforce prohibitions against the slave-trade (Mulder 2022, p. 15) Austria employed them again in closing their markets to Serbian products following the 'Pig War' between 1906-1908, and in 1933 the British government retaliated against the Soviet Union's arrest and trial of six British engineers by suspending commercial negotiations and banning Russian imports. Sanctions have become an increasingly popular instrument of warfare, subsumed under the umbrella of 'economic statecraft' (Baldwin 2020). By the end of 2023, 45 countries imposed sanctions on Russia in response to Putin's invasion of Ukraine in February of 2022. The sheer scale of this regime is not unprecedented, but it is certainly one of the largest economic statecraft campaigns launched at a country since Saddam Hussein's 1990 invasion of Kuwait.[2]

This latest episode of sanctions has involved a new phenomena; the proliferation of large-scale private sanctions that coincided with the imposition of state-initiated sanctions. Private sanctions are sanctions imposed by companies voluntarily by going beyond the compliance requirements of state-initiated sanctions (O. Hart, Thesmar, and Zingales 2023). The issuing of private sanctions began following a large ground-swell of public commentary on who was funding Russia's war, with citizens re-orientating their focus from governments to private actors, as pictures of visible brands in operation across Moscow and St. Petersburg flooded social media. In response, companies began to issue statements on

[1] Quoted from Mulder (2022, p. 1), original source from Wilson (1923, p. 67).

[2] The effectiveness of sanctions is up for debate. Though the sanctions have crippled the Russian economy, and hampered its prospects of growth, they have by no means led to an economic collapse, nor have they curtailed Moscow's aggression.

their position.

Some companies acted swiftly. Just one week after the invasion, Adidas suspended their partnership with the Russian Football Union, and by March 14th 2022, all Adidas retail stores in Russia had closed. British Petroleum (BP), remarked on its website under the heading, 'BP's position in Russia' that 3 days following the invasion, on February 27th 2022, "BP would exit its 19.75% shareholding in Rosneft and other business in Russia" and most remarkably, the company commented, "BP was the first international company to announce such an exit. That decision remains unchanged and BP has no intention of returning to business as usual in Russia" (Petroleum 2022). BP's statement made clear that it was the first to leave Russia of its own violation, not in reaction to sanctions, and the company would follow through on this move regardless of what would come. The company named and shamed Russia, repeatedly using the term "attack".

Yet not all companies reacted similarly. Philips, a company operating in the healthcare space, issued a statement refusing to leave the country, citing humanitarian reasons to remain. Procter & Gamble remained as long as they could without violating sanctions. Unilever, a manufacturing company, refused to leave, but did issue a statement on their website that remained up to one year later. The second sentence states: "Our business operations in Ukraine have stopped*" the asterisk at the bottom of the page in very small font stated: "*Please note, Unilever resumed business operations in Ukraine in April 2022".

Companies that prevaricated found their hesitation costly. In an interview in early March with Time Magazine, Goldman Sach's CEO, David Solomon, made

it clear that individual banks should not make the decision to "ostracize Russia" (Simons 2022). "It's not our job" he said, but the company would comply with sanctions. Following the publication of those comments, the resulting public and employee outcry, Solomon issued a statement, referencing - like BP - that Goldman Sachs was "the first major financial institution" to announce intentions to leave Russia, condemned the actions of Russia, and cited the "responsibility" of the company to leave.

In the several months following the initial invasion, over 1,000 companies publicly announced that they would be retrenching operations in Russia in some capacity that went beyond the legal requirement imposed by international sanctions (Sonnenfeld, Tian, et al. 2022; Sonnenfeld, Mylovanov, et al. 2023). But as the anecdotes above suggest, this was not universal. Many of their multinational peers are still operating in Russia, and made the decision to remain. Some businesses have made no public announcement and are carrying out "business as usual", while others have curbed operations just enough to comply with sanctions. This puzzle poses two questions – why did some companies issue statements to exit Russia? And, what explains the variation in how and when these companies issued statements on their retreat?

Focusing on the latter question, I make two arguments. I argue that firms issued statements on withdrawing from Russia to minimize the reputational costs of remaining amid an online backlash. Rather than concern for preserving a history of socially conscious behavior, early withdrawals by peer firms created a 'reputation stealing' effect, and forced within-industry firms to withdraw within close intervals of their competitors. Existing studies on firm reputation often measure reputation through history-dependent ESG scores. I argue that ESG

scores assume citizen awareness and ignore the salience of high-scrutiny situations. Instead, I suggest that reputational costs are generated by intra-industry diffusion. When one firm signals their intention to withdraw, it increased consumer and public scrutiny of competing firms within that sector, moving the benchmark of 'good' and 'bad' behavior among consumer audiences. The result is a strategic interaction between firms within an industry, where early moving firms publicly signaled an intention to withdraw, and other firms were pressured to follow suit. Firms were forced to move sooner than they anticipated, and publicly disclose their position.

Second, on how firms left, I argue that firms attempted to manage their reputations for future behavior by either tying their hands or using 'cheap talk'. I collected novel data on the content and time points of company-issued statements, including multiple statements from the same company as these firms shifted positions on Russia over time. I show that both remaining and withdrawing firms were sensitive to reputational costs, however remaining firms attempted to appease both international consumers and Russia audiences, while withdrawing firms tied their hands and cut ties with Russia. These types of firms employed specific strategies to maintain reputations with their audiences. Withdrawing firms issued statements quickly, were highly critical of Russia, and chose to publicize their withdrawal using social media. In this way they burnt a bridge with Russia and tied their hands. These firms were unlikely to reverse course and issue additional statements. In contrast, remaining firms were torn between appeasing their international public audience, and mollifying key industrial and political institutions in Russia. These firms face high reputational costs but attempted to balance their reputations among both audiences by focusing attention on their humanitarian activities, leveraging legitimacy through their membership to UN-

adjacent multi-stakeholder initiates, and employing less critical language, i.e. making few references to Russia, leaning on 'compliance' rather than intentional behavior, and avoid terms like 'war' or 'invasion'.

To estimate the effect of reputational concerns on firm decisions to issue statements on Russia, I utilize a database on firm statements to withdraw. I rely on the authoritative list of firms collected by (Sonnenfeld, Tian, et al. 2022) and the Kiev Project (2023) with over 1,000 firms across 62 countries. I supplement this data with an original collection of company-issued statements. For each firm I identify whether the firm issued a statement, the date in which the firm issued a public statement, whether or not they issued a social media statement, and whether or not they reversed this position in the future (either towards or away from withdrawal). I utilized classifications on withdrawal along a spectrum with five grades following the Yale classification system: A) Withdrawal, B) Scaling Back, C) Suspending Operations, D) Buying Time, and F) Digging in. I replicate the methodology as firm's change positions over time - moving forward or backwards along the spectrum. I use financial databases (Refinitiv and Orbis) to collect control variables on these firms. Using Cox proportional hazard models, and logit regression models with fixed-effects, I show that firms face relative reputational costs depending on the actions of their within-industry peers. Reputational costs increase when competing firms signal intentions to leave Russia, causing a 'reputational cascade' effect that motivates their peers to publicly signal their intention towards withdrawing by issuing a statement. Furthermore, I find there is a time-inconsistency effect, and firms are more likely to reverse course as the reputational penalties of remaining in Russia ease when global attention wanes. This suggests the effects of reputation depend on visibility and audience attention. Lastly, I test whether firms that withdraw are more likely

to 'signal boost' their withdrawal to wider audiences. I test competing explanations for factors that could have influenced firm decisions to withdraw at points in time including executive-compensation, shareholder penalties, risk aversion and informational cues.

In the second half of this paper, I utilize computational text analysis techniques to explore the strategies remaining and withdrawing firms used to manage their reputations in issuing statements on their position vis-a-vi Russia. Using word-embedding (GloVe) and topic-modeling, I analyze the content of company statements to better understand how firms communicated with the public. I sourced statement data on surveyed firms conducted by the Business and Human Rights Resource Center (BHRC), a semi-private regulatory body, and manually collected hundreds of additional firm statements randomly sampled from my existing database. The BHRC asked companies to provide answers related to their decision to remain or leave Russia. The findings on cosine similarity suggest that there is a mimicking (or copy-and-paste) effect for statements issued by firms within the same industry, validating the likelihood of intra-industry diffusion and reputational cascades. The results also show that remaining firms are less critical of Russia, and more likely to focus on humanitarian actions and aid. I show that a large proportion of these companies lean heavily on citing both participation in UN bodies (the United Nations Global Compact) and UN guidelines when explaining their decision to stay or leave Russia. Remaining firms leverage legitimacy through affiliation with UN bodies to bolster their reputation, even if they have not left Russia. As a supplementary test, by examining United Nations Global Compact (UNGC) membership data, I find that firms that remained in Russia are more likely than firms that withdrew to join the UNGC post-invasion. These findings suggests that bodies like the UNGC may just be shielding firms

from criticism, and remaining firms learn from peers in their use of IO-adjacent organizations to leverage legitimacy.

This paper makes two theoretical and empirical contributions. Empirically, I construct a novel longitudinal database on company positions on Russia with multiple time-points. This extends existing databases that have currently tracked company positions on Russia in general. For instance, while the (Sonnenfeld, Mylovanov, et al. 2023) database depicts the ultimate position of a company, I tracked changes in company positions day-to-day during 2022. In addition, collecting multiple company statements on each position change allows researchers to examine how companies strategically responded to the war over time. The database adds a measure of company transparency on their position; whether the company issued a statement on a social media platform. This data is impossible to scrape since company statements were often issued using graphics on social media platforms.

Second, I add to the literature on the increasingly prominent role that multinational corporations are playing in international governance and geopolitical conflict. Understanding the patterns of firm behavior during the Ukrainian invasion provides a useful blueprint for future conflicts and how firms align to their home-state positions, as well as how they deal with conflicting consumer and stakeholder audiences. The findings in this paper suggest that while early-moving firms can induce reputational costs on peer firms, and incentivize them to change their behavior even when it is costly to do so, ultimately the effectiveness of reputational pressure weakens over time. This can help inform how effective the use of sanctions and private sanctions can be to pressure state behavior over time. If private sanctions are a critical ingredient of modern warfare,

as they increasingly appear to be, then the preferences and motivations of firms acquires geopolitical relevance (O. D. Hart, Thesmar, and Zingales 2022). Secondly, the findings here shed light on motivating private actors in other complex problem areas of IR where getting firms to voluntarily adjust their behavior is often difficult, such as climate change.

Lastly, this paper offers two new perspectives on how reputation is both measured and theorized in IR. The current literature on the role of reputation as a mechanism for inducing firm behavior often uses ESG as a proxy for reputation. I argue that ESG is a poor measurement for firm reputation. ESG treats reputation as a historical concept where firms build social capital over time, and that stakeholders and consumer audience observe and catalog firms. The finding here suggest that reputation can be re-conceptualized as a group-based phenomenon, where observer judgments over reputation depend on the average baseline of the group. When an early-mover departs from the baseline, observer judgments over individuals in the group begin to shift. Related to this, rather than historical, reputations here are situationally determined. High-salience situations create higher costs and increase scrutiny.

Extant research on reputation in international relations often assumes that reputational costs are the independent variable that affects how agents make decisions. In this paper, I attempt to show how firms strategically manage their reputation when they anticipate fall-out. Many accounts of reputation in IR suggest actors are more or less sensitive to reputational costs, while here I attempt to show that agents face competing reputations among multiple audiences. I contribute to a growing body of literature that evaluates the role of MNCs as instrumental actors in international relations, who play a role in transnational

issues independent of states (Thrall 2021; Miura and Kurusu 2015; Amer 2018; Rasche, Gwozdz, et al. 2022).

## 2 Sanctions and Firm Reactions

Private sanctions are an increasingly important component of contemporary warfare, with the withdrawal of over 1,000 companies in 2022 following Russia's invasion of Ukraine (Sonnenfeld, Tian, et al. 2022). The logic of private sanctions departs sharply from that of state-imposed sanctions. Firms must oblige with many of these state-sponsored sanctions, for instance when the US Treasury Department imposed sanctions on Russia's largest banks, foreign and domestic banks and corporations were forced to curtail financial interactions with these sanctioned institutions. And price-caps on Russian oil through secondary sanctions meant punishment for foreign buyers who would be prevented from conducting businesses with American companies.

However, voluntary curtailment of operations that goes beyond the legal requirement set by international sanctions at this scale is unprecedented (ibid.). Although recent, a growing literature has sought to explain this phenomenon with mixed results. For instance, one study points to the composition of shareholders of companies (O. Hart, Thesmar, and Zingales 2023), with shareholders who score high on compassion and authority opting to withdraw, however this effect seems very much curtailed by the potential monetary costs of withdrawal. In contrast, using interviews with Japanese business executive, Davis, Jialu Li, and Miyano 2023 argue that a strategic environment creates cues that shape managerial perspectives by learning about potential withdrawals from peers. The authors, however, find no evidence for reputation motivating behavior. In

another study, Lu, Huang, and S. Li 2022 finds evidence for several channels that affect firm decisions to withdraw, including social awareness and consciousness, the decision of a domestic government of which the firm is headquartered to sanction Russia, institutional quality and common law of the home state, and reputation. More recent work has added interesting insight into tracking which firms exited Russia or changed ownership (Wellhausen and Zhu 2024), finding that consumer-oriented firms, already under Russian management, and those not operating within Russian industries were most likely to exit. The issue of tracking true company exit is difficult, with the Wellhausen and Zhu (ibid.) suggesting that only, 33.3% of firms withdrew. This paper does not discuss which firm have truly exited, and is instead concerned with whether companies issued statements signaling their withdrawal, or how companies communicated to the public on their position, rather than true or successful Russian withdrawals given the time it takes for a company to exit, or companies which were taken over by Russian administrators.

This growing literature promises insight into the unprecedented imposition of private sanctions, however the findings remain contested on explaining why firms issue statements to withdraw, and exactly how this reputational mechanism is influencing firm behavior. This theoretical gap has left many questions unanswered. For instance, is it the case that firms that remained in Russia were not impacted at all by concern for losing social capital among a politically-charged consumer base DellaVigna, List, and Malmendier 2012; Ottoni-Wilhelm, Vesterlund, and Xie 2017; Lyon and A. W. Montgomery 2013; Krieg and John Li 2021? Is there an early-mover advantage to issuing statements on withdrawing, and if so, does the timing of firm decisions to withdraw matter? What explains the shift of firms from early decisions to remain in Russia and weather out the geopo-

litical storm, to suddenly departing? And why have some firms renegaded on their initial decisions to depart? For instance, WeWork, who made promises to leave in March 2022, reversed course soon after with options to book co-working spaces in Moscow as of 2023, or Nestle who initially promised to only sell essential goods, but continued to provide chocolate bars. If reputation plays a role in answering these questions, there is more nuance to the mechanism.

While interviews with shareholders and executives, and leveraging stock-data and stock-price returns has provided interesting insights into the facets of internal company decision-making, limited data on the strategic movement of firms, and the nature in which these companies left Russia, has hamstrung research on reputation and private sanctions. By leveraging data on the strategic timing of company withdrawals, reversal of decisions, and how firms signaled (or did not signal) their position on Russia, the role of reputation and the mechanism through which it effected behavior becomes clearer, for motivating both withdrawals *and* remainders.

## 3 A Theory of Reputation and Intra-Industry Diffusion

The early imposition of private sanctions and firm withdrawals following Russia's invasion revealed an emerging pattern. When an early moving company issued a statement intending to withdraw or scale back operations in Russia, competing firms within that sector would often follow suit, in some cases within hours. Take for instance the reactions of holding companies with presence in Russia. Within one week, successive holding firms announced withdrawals from Russia. Accenture announced their exit first, followed by WPP and Publicis. Havas quickly issued a statement the day after, announcing it had frozen new

investments in Russia. IPG followed the next week, with Dentsu quickly after. WPP announced its decision to discontinue operations in Russia March 4, 2022. Within 10 days, IPG announced it would suspend operations in Russia.

Employees in these companies pressed executives to withdraw or scale back operations. In particular, Publicis Groupe had executives questioning in company meetings why the firm remained in Russia, arguing against generating revenue from ad buys in Russia, and how this would benefit Putin's war effort. Despite this pressure the company did not issue a statement. In contrast, IPG executives pressured the company to issue a statement of withdrawal (Coffee and Rittenhouse 2022). Publicis had the largest operation amongst holding companies in Russia, with around 1,000 employees at a wholly-owned division of agencies (Starcom and Digitas), as well as Russia's largest state-owned digital services provider as a client (ibid.). The company was pressed for comment following the announcements of its peers, and just two days after its main competitor, IPG, voiced intentions to exit, Publicis announced their withdrawal on March 16, 2022.

The pattern continued with other holding companies. Omnicom, a holding company with Kremlin-owned clients, including the Russia Agricultural Bank and Rosbank, announced its intention to withdraw on March 17th 2022. French holding company, Havas, who held affiliates with Russia, announced suspension of operations on March 10th. Japan's Dentsu had a strong presence in Russia, with Russian-based offices for multiple agencies including Carat, iProspect, and Isobar. Additionally, Dentsu had launched a new Russian division in 2020 as a joint venture with OKS group (ibid.). Despite all this, following peer with-

drawals, Dentsu announced their withdrawal from Russia on March 16th.

The pattern is revealing; firms move with their peers, regardless of their country of headquarters. Company statements which signaled withdrawals, reinforced this observation. Early-moving companies would proclaim themselves to be the first to withdraw within their sector (rather than a follower firm, subject to public pressure to move). These early-movers highlighted that their actions either preceded the imposition of sanctions, or that their actions went above and beyond the requirement of sanctions. Take for instance, Lithuanian oil company, Orlen Lietuva, who published a statement on their social media Linkedin account. The statement read:

> "The first in Europe and THE ONLY ONE as soon as the war in Ukraine began, without waiting for the sanctions that refused Russian crude oil... The first in Europe and the ONLY which turned away from Russian crude oil even before the introduction of sanctions. These 10 points prove that we have solid grounds to say that our role in the economy and energy security of Lithuania is of particular importance."[3]

In another post, the company went on to say:

> "We had to reorient ourselves, without waiting for the sanctions to come into force, we abandoned the import of Russian oil raw materials, the logistics chains for cargo transportation to Ukraine changed... And AB ORLEN Lietuva with the only crude oil refinery in the Baltic States is no exception. We had to make a shift, we shun imports of Russian crude without waiting for sanctions to enter into force, and made changes in

[3] (Lietuva 2022)

the logistics chain of the product shipments to Ukraine."

Some companies went further by severing relations with their trade associations that failed to act – effectively working to 'sanction' membership, and to avoid any negative reputational externalities of associating with companies that refused to denounce Russia. Take Pandaora for instance, the company suspended operations on the 24th of February 2022 – an early mover in contrast with its competitors – and issued a withdrawal statement on the 30th of March 2022. That same day, the company issued a statement announcing it had severed its "12-year membership with the trade association Responsible Jewelry Council, RJC" for "failure to suspend Russian companies' memberships and responsible business certifications" and urged other members to suspend business with Russia. Pandora had previously requested that RJC take such actions, CEO Alexander Lacik remarked:

> "We are shocked and saddened by the unprovoked attack on Ukraine, and our thoughts go to the people of Ukraine, who are victims of this senseless act of military aggression. The war requires all businesses to act with the utmost responsibility regarding any interactions or business dealings with Russia and Belarus. Pandora cannot in good faith be a member of an association that does not share our values."

Pandora's competitors followed their lead. Richemont and Kering exited RJC within 24 hours. ALROSA voted to suspend its membership with RJC within 48 hours. Yet, in the weeks that followed, when shrouded in the shadows, these companies returned. In response to this small scandal and showmanship, RJC publicly remarked, "Most of the small number of members who briefly left the RJC in late March rescinded their resignations within a few weeks, including all

the Kering and Richemont Maisons, retaining their long-standing RJC membership".

These examples illustrate a pattern of behavior of withdrawing firms that suggest two phenomenon. First, that firm decisions to issue withdrawal statements from Russia were influenced by reputational costs induced by peer behavior. And second, that this reputational mechanism was temporally constrained; as audience attention weakened, incentives to withdraw also weakened.

The problem with identifying whether firm's were influenced by their concern for reputational costs is that firm reputation is usually measured through a history-dependent proxy, rather than a group-based measure that takes into account the behavior of a firm's competitors. Prior work has linked firm ESG and Corporate Social Responsibility (CSR) to financial and/or stakeholder concerns over investors and market risks, or reputation to endogenous historical ESG ratings over time. These types of arguments have been applied both explicitly to the Russia-Ukraine war, but also to firm behavior over carbon disclosures or other high-risk events. Past work has used ESG scores to proxy for a firm reputation, since firms with higher ESG scores have higher social capital at stake, greater social consciousness, and more internal and external pressure to follow social norms (Lu, Huang, and S. Li 2022).[4] There are two main problems with using past ESG outcomes as an indicator of firm reputation. The first is that the measure groups firm outcomes on social, environmental and governance issues into a single rating. ESG scores can offset bad behavior in one area, with good behavior in another, and usually involve lower-stakes issues, or issues with at least some certainty. Second, for a corporate actor to suffer reputational costs amongst a consumer audience, two criteria must be met; (1) the event must be

[4] See Appendix for more on the use of ESG as a proxy for reputation.

visible and highly salient, and (2) audiences need relevant benchmarks to make easy judgments on that actor's behavior. Even if a decomposed measure of ESG is used (i.e. Governance or Social measures), they're constructed with information that audiences may be unaware of.

ESG scores become a problem for the first criteria since they are tabulated across a range of firm behavior and actions over the course of year(s), much of which is largely invisible to consumers. For ESG ratings to matter to consumers, the company's behavior must be visible enough to be picked up by mainstream media. Existing studies suggest that most news that affects company ESG ratings *and* in turn affects consumer ratings of companies involves large events and poor company behavior (rather than positive behavior) in that context (Meier et al. 2023). The Ukrainian-Russian war constitutes an unprecedented crisis, that is highly visible, costly, and involved multiple consumer boycotts documented on online social media platforms. The costs of corporate action is not just constrained to the possibility of firms that leave or divest being unable to recoup their divestment's or return to the Russian market; strong withdrawals from private actors can send a signal to other politically-unaligned markets that these firms are not neutral actors, and may behave similarly in future conflicts. Firms can also avoid negative ESG ratings by merely suspending operations in Russia and making humanitarian aid donations to NGOs operating in Ukraine.

Second, audiences need some type of benchmark in which to make judgments on 'good' versus 'bad' behavior. As Jervis notes, "the burdens of calculation are further reduced by the use of benchmarks to guide decisions" (Jervis 2017, p. 44). Benchmarks are useful when individuals are mobilizing against firm action online, and such benchmarks can provide a way for individuals to coordinate be-

havior and guide collective action (Schelling 1980, ch.3). That benchmark can be generated through either comparing current behavior against previous behavior, or by comparing alike actors.[5] While we might suggest that a firm's behavior during the Russian annexation of Crimea in 2014 could be a comparative situation in which to benchmark a firm's behavior, the event is temporally too distant for the average consumer to remember individual firm reactions. Consumers are not comparing apples with oranges (or, energy companies with clothing retail stores). Treating all firms from all sectors with a high-ESG rating as having 'the most social capital to lose' ignores the fact that consumers are more likely to be primed to pay attention to a firm when its competitor makes a costly move to withdraw. Sub-sectors become the dimensional trait within which the consumer can establish a benchmark (Slovic and MacPhillamy 1974; Slovic 1975). While the external calculation of a firm's cumulative ESG rating might be a useful indicator of social consciousness for within-industry elites or qualified investors, the average consumer is more likely to use an easily available comparison, such as Subway's decision to remain in light of McDonald's decision to leave, this provides an easy heuristic for who is a 'good' or 'bad' types for a firm's human rights reputation (Kahneman and Tversky 1973).

The process I describe occurs when one firm within a sector moves first in reaction to the initial shock. In this case the shock is Russia's invasion of Ukraine. As early mover firms within a sub-sector shift their behavior and actions, this in turn causes consumers to adjust how they perceive competing firms who have chosen to remain. Consumer audiences survey firms and make judgments on the relative 'good' and 'bad' actors based on their peers. There is evidence in the literature that supports this theory; in markets with network

[5] Or, importantly, the perception that two things are alike even if this may not be true.

effects (M. L. Katz and Shapiro 1994), where first-and-early-movers enjoy advantages over later movers, although this research is primarily concerned with firm survivability (Kauffman and B. Wang 2007), stock market capitalization and revenue, and securing a customer base (Lieberman and D. B. Montgomery 1998). The same logic should still apply to other contexts where diffusion effects are likely and the issue is high-salience.

When firms within industries seek to enhance, protect or maintain their reputation it can become a positional good. The firm that choose to withdraw may feel better and look better amongst their consumer base, but it can hurt the firm's neighbours (within that industry) by making that competing firm with the closest proximity look worse, as Bénabou and Tirole 2010 terms, a "reputation stealing externality". As participation in withdrawals from Russia increase, early movers may be rewarded for showing altruistic behavior, while those who have abstained from issuing a position, suffer a social stigma. When withdrawal rates increase and participation of firms in withdrawals increases, the benefits of leaving decrease (since the behavior has become more common), but the stigma of not leaving increases (as the spotlight on the very few who chose not to leave becomes more concentrated and less diffuse amongst many actors). This effect is not new and has been demonstrated in both individuals and organizations (Bénabou and Tirole 2010; Monin, Sawyer, and Marquez 2008). This effect would suggest that for members of a sub-sector (their group), concern for reputation after an event will be time-dependent. Early-movers and those who react quickly improve their reputation or prevent losses. Over time, the firms that chose to remain are less likely to reap these reputational benefits, or repair reputational damage the company incurred for lagging behind, making it less likely these firms are influenced by their peers. Announcements of an eventual with-

drawal could remind the public of how long it took this company to withdraw. I hypothesize that this logic would suggest the following effect of peer-reputation on firm withdrawals:

> Hypothesis 1a: *Industry-diffusion Effect.* As the issuing of withdrawal statements of firms within an industry increases, member firms of that industry will be more likely to withdraw at a quicker rate.
>
> Hypothesis 1b: The industry-diffusion effect on firm statements will weaken over time.

For the reputational mechanism to be at play, there must be some initial public pressure exerted on the company. The salience of the Russian invasion, combined with the international coverage of the event, increased the reputational costs for companies to issue statemnts on remaining. The role of media coverage on increasing reputational costs is supported by past findings in the literature. Dyck, Volchkova, and Zingales (2008) for instance focused on media coverage of corporate governance in Russia between 1999-2002, a period of common and highly visible corporate governance violations, with ineffective and absent grievance mechanisms.[6] The authors found that news coverage by Anglo-American newspapers, not local press, led to reversals of bad corporate governance decisions. The mechanism here is that the effect of coverage in exposing bad behavior and shaming companies is only effective if the consumer audience shares the same values. Media coverage works through the mechanism of increasing the reputational costs of bad behavior via a relevant audience that shares the same values. If the effect of this reputational mechanism in inducing firm good behavior (and reversal of bad behavior) was visible in 2002 when mo-

[6] No derivative suits at the time, and courts were easily corruptible, for more see: Slinko, Yakovlev, and Zhuravskaya 2005.

bilization of collective action was more costly, and public backlash more diffuse, then the effect should be much stronger in a 2022 setting where online platforms increase the visibility of brand behavior (Jansen et al. 2009; Papacharissi 2015; Endres and Panagopoulos 2017).[7]. There are many examples of this type of 'peer pressure' induced reputational cascade effect on corporations and issuing statements. When one company acts, it creates a reputational externality for the companies that fail to act, by moving the relative reputational benchmark. Take for instance the PRISM-Wikileaks scandal, companies initially hoped to stick to 'no comment' but "a version of peer pressure kicked in, said one chief executive who spoke on the condition of anonymity because of the fragility of the situation. Once one company issued a flat denial, the others felt they had to follow suit" (Sorkin 2013).

One way to identify variation in public pressure on firms would be to track social media boycotts on individual firms [8]. However, changes to academic access to platforms like X in 2022 has made this type of data more difficult to obtain. In the absense of this information, a common way to proxy for social awareness, informational updating on firm behavior, and the amount of pressure individuals could exert on a firm, is to use internet freedom and internet subscription rates across countries (Lu, Huang, and S. Li 2022; Durante and Zhuravskaya 2018; Chen and Yang 2019; Manacorda and Tesei 2020; Enikolopov, Makarin, and Petrova 2020; Guriev, Melnikov, and Zhuravskaya 2021). Firms headquartered in states with higher rates of internet freedom and internet subscriptions should be more likely to withdraw from Russia. This finding would support the

[7] A recent and interesting demonstration of stakeholder pressure on corporations to issue 'targeted' statements can be found in Z. Li and Disalvo 2023 work on the influence of partisan stakeholders on major US company decisions to withhold PAC contributions following the 2021 Capitol insurrection

[8] See for example Pajuste and Toniolo (2022)

reputational mechanism.[9]

In addition, companies had some control over the visibility of their actions, and the nature of the signal they sent. Companies largely responded by either issuing a statement on their website, or issuing a statement through their public-facing social networking platforms of which they have sizeable followers (Twitter(X), Linkedin, Instagram, or Facebook), and vast reach. There is a substantive difference between website statements issued by firms, and social media statements issued by firms. Although both send a signal, social media statements can be considered 'signal-boosters'. Effectively they take the firms message issued on their website, and amplify it to a broader audience.[10] While almost every company that issued a social media statement *also* issued a website statement, not all companies opted to send social media statements.[11] Of course, some firms chose to issue no statement, and while a firm issuing a statement on the issue of Russia is still a substantive signal – conceptually, social media signals boost signals if we consider the proximity of the message to the consumer. Consumers are more likely to find and be exposed to social media statements than website statements. [12]

I argue that early withdrawing firms will be more likely to boost their signal by issuing statements outlining their position on social media platforms. These platforms provide direct linkages between firms and their core audience.

[9] See appendix for more details on this.

[10] Another term for this could be targeted and non-targeted responses, as used by Z. Li and Disalvo (2023), however targeted would more aptly refer to the directness of the signal, for instance companies that issued 'withdrawal' statements versus non-targeted which issues 'suspension' statements, than the medium of the signal.

[11] 50.4% of the firms in my sample issued a social media statement. While 84.9% of firms in the sample issued a website statement. Of the firms that issued a social media statement, 95% also issued a website statement.

[12] Some firms have deleted their company statements since issuing. For example, Waters Corp and Krispy Kreme have both deleted their website statement since 2022.

In contrast, I argue that firms that remain in Russia are attempting to balance two audiences at once – i) their consumer audience and ii) Russian audiences. Remaining firms are seeking to avoid the reputational costs of remaining while avoiding the reputational costs of suspending operations to comply with sanctions. These firms will be less likely to issue social media statements and less likely to seek out means to boost any signal. Firms that are in a purgatory limbo state (one foot in, one foot out), will also avoid issuing social media statements. This finding would lend stronger support to the role of reputation in inducing firm's to issue withdrawal statements.

> Hypothesis 1c: Firms that withdraw will be more likely to issue social media statements than firms that remained.

As I have suggested above, firms that remained in Russia were no less concerned about the reputational costs of remaining. These firms instead attempted to balance the reputational trade-offs between two separate audiences; i) Russia (regulators, authorities, and consumers)[13], and ii) their international consumer base. The firms that withdrew from Russia early issued strong statements that were highly critical of Russian actions. In doing so, these firms effectively burnt the proverbial bridge of continuing business in Russia in the future. Their actions tied their hands, and sent a strong signal to their international consumer base that they were committed to following-through on withdrawal.

Reputational concerns affect the nature of firm decisions to issue statements to withdraw or remain, but they also influence the nature in which these companies signal their decision to their audiences. Remaining firms were deeply concerned with the increased scrutiny over their decision to remain, and the effect on

[13] An argument could be made that included in this group is other authoritarian countries who prefer politically-neutral firms

their own reputation as a result of peer withdrawals. For instance, Limagrain, a French international agricultural co-operative, decided to remain in Russia and cited that sanctions did not apply to the company. However, the company established a crisis task-force dedicated to managing "the risks of damage to Limagrain's reputation or image as effectively as possible" (Limagrain CSR Report, 2022, pg.26).

Remaining firms attempted to balance multiple audience demands and preserve conflicting reputations amongst these audiences. Protecting the company's credibility amongst its international consumer base may require the firm to present itself as a 'good' type by trying to either deflect from the firm's continued presence in the region – issuing statements claiming the firm is donating to humanitarian international organizations and NGOs – or to justify the continued presence by making claims that the firm has a "moral responsibility" to their employees in Russia, or that (in the case of many pharmaceutical firms), it would be amoral to punish Russian citizens by withdrawing. On the other hand, the firm's continued functioning in politically-contentious regions in the future may require the company to project a 'political neutral' image to its other audience (Russian regulators, the Kremlin, or even China). The firm achieves this in a few ways. First, the firm issues statements outlining that the firm's suspension of operations are the direct result of a legal necessity to comply with sanctions, rather than out of the volition of the firm, or the firm's judgement. The firm avoids any politically-charged language, in some cases this means an avoidance of references to Russian entirely, where the statement is exclusively focused on humanitarian relief in Ukraine and compliance with "relevant sanctions". In more explicit ways, this means the firm avoids referencing the word "invasion" and in some cases "war". This follows Kremlin directives framing the situation

as a "special military operation on the territory of Ukraine" and Putin signing a law in March 2022 that criminalized reporting that opposed the war against Ukraine, with critical labels like 'invasion' and 'war' labelled as undermining the national interest, and constituting 'false information'.

> Hypothesis 2a: Remaining firms are less likely than leaving firms to use critical language towards Russia, and more likely to focus their attention on humanitarian aid.

The second strategy these remaining firms used to protect their reputation amongst conflicting audiences was to leverage the legitimacy of human-rights adjacent international institutions.

## 4 Multi-Stakeholder Initiatives: Reputation-Insulation

Firms that choose to remain in Russia anticipated suffering reputational costs, and as such, these companies employed a variety of strategies to deflect, hide, or mitigate consumer backlash from international audiences. One such way that companies employed this strategy was by leveraging legitimacy through UN-adjacent voluntary regulatory bodies. Firms advertised their membership, or association, with these human-rights institutions to obfuscate their continued business practices in Russia. Two key global public-private initiatives used were the United Nations Global Compact (UNGC) and the United Nations Guiding Principles (UNGP).

Debates over the role of public-private voluntary regulatory initiatives like the UNGC are split between critics who argue that these organizations function as 'wrapping paper' for firm practices across social, environmental, and governance issues, allowing corporations to leverage legitimacy without follow-through

(Miura and Kurusu 2015; Rasche, Waddock, and McIntosh 2013; Bruno and Karliner 2002), and proponents of these organizations who argue that these bodies can supplement ineffectual domestic regulations that encourage learning to socialize members into changing their behavior, or by fostering collective action in the private sector on transnational issues (Rasche 2009). Research explaining patterns of corporate social responsibility adoption amongst corporations has shown that firms with greater international linkages to governmental and non-governmental organizations, combined with consumer receptivity to CSR principles, influences the ceremonial, but not substantive, adoption of CSR frameworks (Lim and Tsutsui 2012).

An example of this is the conduct of one of the remaining companies in Russia, Robert Bosch GmbH, a German multinational engineering and technology company. Ironically Bosch's director of migration praised the company's membership to the UNGC as being responsible for the quick movement of corporations in responding to the Russian invasion, even as Bosch has one of the worst failing scores on Russian action ('Buying Time') in 2022. The company continued to make references to its membership to the UNGC despite its continued presence in Russia; "International agreements such as the Global Compact for Migration make it possible for actors to respond quickly and concertedly in moments of crisis, as Europe did after Russia went to war against Ukraine", stated Raphaela Schweiger, the Director of the Migration Program at the Robert Bosch Stiftung

### 4.1 UNGC: Background & Membership

The United Nations Global Compact (UNGC) describes itself as the "world's largest voluntary corporate citizenship initiative" (UN Global Compact Office

2011). The UNGC transforms a burden of responsibility that was once the exclusive domain of states to the private sector. The UNGC's mission is"attempting to ensure that economic and social interactions are respectful of human rights as well as socially and environmentally sustainable" (Bernhagen, Mitchell, and Thissen-Smits 2013).

Although joining the UNGC is voluntary, the organization is an example of firm participation in a norms-based responsible governance initiative. Joining and maintaining membership to the UNGC requires overcoming some entry barriers, with costly updates. First, chief executives must seek and obtain board support to join, and make annual financial contributions to the UNGC.[14] In order for companies to retain their 'active' status within the organization, they must submit annual reports outlining how the company is making progress in four areas; human rights, labour, environment, and anti-corruption. Much of the past literature on the public-private voluntary regulatory initiatives has focused on the environmental division of the organization (Lerner and Osgood 2023; Nilsson and Fougère 2024; Coen, Herman, and Pegram 2022). Although some studies have examined the UNGC in the context of human rights (Thrall 2021; Schembera 2018). Notably the organization makes clear that good practice in one area does not offset harm in another.

For members to maintain their 'active' status within the organization, they must pay an annual fee, and each member must submit an annual Communication on Progress (CoP) report that details how the firm is making progress on these four areas; failure to submit a report would lead to the company being assigned 'non-communicating' status, and eventually 'delisted'. These status

[14] Annual contributions cap out at $30,000 USD conditional on member location and relative revenue.

updates are publicly disclosed and remain on the website. Extant research has reached mixed conclusions on the effectiveness of the UNGC in promoting better ESG practices. Some studies suggest principle-based Public-Private Partnership initiatives like the UNGC "suffer from lower levels of output legitimacy as they do not monitor participants" (Mena and Palazzo 2012; Sethi and Schepers 2014), however the CoP reporting requirement is meant to countervail the visibly low effect of membership to the UNGC on legitimate CSR changes (Hamid and Johner 2010). Other studies have found that investor's penalize firms for falling into non-communicative or delisting status (Amer 2018), suggesting the organization has an indirect enforcement mechanism and may influence firm behavior.

The UNGC is an ideal case for investigating the role of semi-private regulatory bodies in providing reputational cover for firms, and influencing firm behavior to issue statements withdrawing from Russia, for three reasons. First, membership is broad enough to cover different sectors,[15] geographical locations, industries, and firm sizes, allowing me to test the effect of membership on a large sample of multinational firms involved in the Russian exodus. Second, the UNGC has low entry costs with no explicit targets, no independent monitoring capacity, and offers the possibility of redemption (moving from 'delisted' to 'active' after submitting CoPs), but leverages its legitimacy through its affiliation with the United Nations. This means smaller firms with resource-constraints do not face enormous barriers to entry. Third, the UNGC is not just loosely affiliated with the situation in Russia through its general guiding principles on responsible business practices; the organization issued explicit guidance for member firms to curtail their business presence in Russia to "make sure that they are not complicit in human rights abuses", with the direct message to members that "the

[15] With the exception of tobacco manufacturers, and companies manufacturing and producing landmines and cluster bombs, who are prohibited from joining the UNGC.

lines between politics and economics become increasingly blurred, the question is no longer *if* responsible businesses should engage, but *how* they should do so"(Ojiambo 2022).

### 4.2 UNGP: Background & Membership

The second IO-adjacent regulatory institution for firms that has relevance to the Russian-Ukrainian war is the the United Nations Guiding Principles (UNGP). Proposed by John Ruggie in 2008 to "anchor the business and human rights debate".[16] Established by the United Nations in 2011, the guiding principles were accepted with little resistance and principally absorbed by many multinational corporations (Addo 2014), although they were critiqued by certain NGOs for setting too low of a bar (Sanders 2014; Ruggie 2011; Deva and Bilchitz 2013). Through the UNGP Reporting Framework, companies can choose to make disclosures to the UNGP on managing human rights risks. Like the UNGC, the UNGP also set out expectations over corporate behavior during wartime and decisions for private actors to disengage. While the "mere presence in a country, paying taxes, or silence in the face of abuses" did not constitute legal liability, it would constitute a violation of the UNGP" (Ruggie 2008).The UNGP specifies that a company should disengage if it cannot guarantee adverse human rights impacts (*Responsible business conduct implications of Russia's invasion of Ukraine* 2023).

[16] John Ruggie, Report of the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises: 'Protect, Respect and Remedy: a Framework for Business and Human Rights'.

### 4.3 MSI's as Reputational Management

To test whether membership to these institutions affected firm's issuing statements to withdrawal, and encouraged proactive statements of withdrawal, or alternatively if these organization provided an easy 'blue-washing' mechanism, I utilise firm membership data and collect firm statements on the company's position on Russia as text-as-data. I make two arguments regarding how remaining firms utilized the UNGC to protect their reputation. First, membership to the UNGC will have no effect on initial firm decisions to issue statements to withdraw. Given the high-stakes nature for some firms of withdrawing, and the uncertainty of consumer-attention on Russia in the future, the likelihood that these initiatives can constrain firm behavior is low. Existing studies so far suggest that these institutions have little impact on constraining firm behavior on human rights, environmental impacts, or other corporate social responsibility issues.

Second, firms who remained in Russia were more likely to make references to multi-stakeholder initiatives and responsible governance initiatives that are adjacent and linked to the United Nations, in order to leverage the legitimacy of these organizations to blue-wash ("wrapping their wrongdoings with the UN blue flag and enhancing their reputation"(Miura and Kurusu 2015, pp. 286–287)). To the extent that these initiatives encourage 'learning', I argue that just as reputational costs fluctuate in reaction to firm behavior, defecting firms learn from their peers on how to mitigate reputational costs by leveraging the legitimacy of voluntary governance initiatives to covertly hide misconduct. Thus, Remainder firms will be more likely to make reference to the UNGC and UNGP in their

official statements justifying their presence in Russia. [17]

Hypothesis 2b: UNGC membership will have no effect on the decision of firms to withdraw from Russia.

Hypothesis 2c: Remaining firms will be more likely to reference humanitarian action and UN-adjacent multi-stakeholder initiatives in their company statements.

## 5 Data

The main data set consists of over 1,300 firms across 68 countries and 13 different industries across the course of 9 months. The list includes firms added by Sonnenfeld, Mylovanov, et al. 2023, and supplemented this list with the KSE database (KSE Institute Project). The KSE database is extremely comprehensive, has more geographic diversity than the Yale list, and includes data number of employees, market exposure, and other important financial indicators.[18] The list includes large US firms that were (or are) significantly exposed in Russia, firms - public and private - of varying size and market share in Europe and Asia. I used these lists to first collect company statements. The dependent variable is a scale of withdrawal with five categories, I used the Yale withdrawal rubric developed by ibid. which grades firms along an A-F letter grade scale that reflects the difficult calculus of curtailing operations in Russia. Each grade is assigned

[17] In the appendix, I test whether remaining firms were more likely than withdrawing firms to have joined the UNGC in the post-invasion period (2022-2023). Past research suggests that late adopters of multi-stakeholder initiatives "are primarily driven by isomorphism and conformity, which can enable decoupling behavior"(Rasche, Gwozdz, et al. 2022; Bromley and Powell 2012). I find preliminary evidence for this, which helps support the theory that remaining firms learnt and adopted similar reputation-protecting strategies like their peers.

[18] For more see: .

through a team of experts and the grade reflects expert consensus. Crucially for the purposes of this paper, the grading is based on company's public statements and actionable evidence of movement towards withdrawal (that is SEC filings on regulatory compliance, financial disclosures, and website statements) rather than private correspondents or confidential material. The rubric follows as:

**A: WITHDRAWAL**: companies making a clean break/permanent exit from Russia or and/or leaving behind no operational footprint.

**B: SUSPENSION**: companies temporarily suspending all or almost all Russian operations without permanently exiting or divesting.

**C: SCALING BACK**: companies suspending a significant portion (but not all) of their business in Russia.

**D: BUYING TIME**: companies pausing new investments/minor operations in Russia but largely continuing substantive business in Russia.

**F: DIGGING IN**: companies defying demands for exit or reduction of activities largely doing business-as-usual.

Since Sonnenfeld, Mylovanov, et al. (2023) grade firms by their final action, I did not follow their grading for specific companies and instead collected data on company initial positions, dates statements were issued, and subsequent movements. From the five grades, I constructed a weighted variable between 0-1 that represents the extent to which the company has withdrawn. 0 indicates the company has reversed its position, and 1 indicates the company has withdrawn. In

addition, I added an additional category to the dataset that doesn't fit into this scale; 'Scaling up' which indicates that the firm's first action in response to the crisis is to expand their operations - many Chinese and Indian firms fit this category. I also added a binary dummy variable 'reversal'; which is when a company reverses its original position and move 'backwards' along the scale, for instance if the firm moves from 'suspension' (B grade) to 'scaling back' (C grade) which might indicate a resumption in activities, or slow reentry into the market. 0 represents no reversal in position, while 1 represents a move backwards. I also construct a conservative dummy `withdrawal` variable that assigns a firm a value of 1 for withdrawing, and 0 otherwise.

It is important to note that these lists of firms do not represent an exhaustive list of all foreign-owned multi-national corporations that exist or did exist in Russia.[19] Crucially, these list contains firms that were explicitly named and shamed online, with the information available to international users, with the list names widely shared on Twitter. The Yale database was built in part by individuals contacting Yale to add businesses to the list, businesses themselves adding themselves to the list, and Yale researchers contacting companies and adding them to the list. Thus, the list already represents a fairly visible group of firms. Similarly with the KSE list, this database of firms is partly built off of Yale's School of Management database, the University of St.Gallen work, epravda.com.ua, squeezingputin.com, https://bloody.energy/ and leaverussia.org websites and other open sources. The KSE authors also created a Telegram bot that tracks and monitors news on priority foreign companies and was highly visible to Telegram users. As a reuslt, the KSE list represents and even larger universe of publicly named-and-shamed firms and their operations online. The data collected by both

[19] See Wellhausen and Zhu 2024 for current research that includes an exhaustive list with fine-grained data on MNCs in Russia.

sources is a better resource for specifically uncovering the effects of reputation on firm withdrawals since the list was highly visible, garnered large media attention, and was used as a visible benchmark for consumers to evaluate alike-companies (A grades versus F grades).

### 5.1 Reputation: Measuring Diffusion Effects

Measuring the effect of reputational externalities induced by early-movers within a sector is difficult. However, if we treat the sub-sectors, and country-level as groups to which firms are identified by consumers as belonging to, we can use this group as a reference point for a cascade or diffusion effect. When one member of the group issues a statement to withdraw, they create reputational externalities on other members of the group. The resulting cascade (or diffusion) will influences firm decisions to issue statements to withdraw.

Firms operate within a sub-sector group; the group itself serves the interest of consumers choosing between alike products; when comparing companies to each other, individuals are more likely to use firm competitors as the reference point. For instance, consider the actions of two unalike companies: A consumer goods shoe and clothing company, Adidas, issues a statement to withdraw from Russia at the same time as an auditing firm also withdraws. Consumers of Adidas are more likely to notice, inquire or seek out information regarding the actions of competing firm Nike, not the auditing firm. The 'early mover', Adidas, creates negative reputational externalities on their competitors as the reference point of 'good behavior' has shifted and non-compliers become more obvious to consumer audiences. In this way, the early mover creates a reputational cascade within their sub-sector group, where competitor firms now move to act to avoid social penalties and reputational damage. Similarly, to sector diffusion, firms

also belong to a country-level group; if early mover firms in a country, such as the United States, issued statements to withdraw or suspend activity in Russia, they will create reputational pressure on their peer firms also headquartered in the United States. In this instance, the audience is the country-level population, and the reputational cascade effect takes place between firms within a country.

Reputational interdependence can have many interpretations, however reputational interdependence ultimately stems from 'optimizing behavior' Cook, Hays, and Franzese 2023, thus a positive effect means that the sector-level group makes firms race towards suspending or withdrawing from Russia, while a negative effect means that sector-level firms deem that their optimal choice is to preserve (or even expand) their revenue and market position while their peers suffer the costs of withdrawal or suspension because they have prioritized the moral or social imperative to leave Russia.

I include time variables (day/month/year) in which the firm issues its first and second statement of action or a statement on their position on the conflict. I manually collected this data on public statements through website platforms, issued statements to the media, and social media posts by the company. If the firm has issued two statements on their website and social media with different dates, I use the press release date from the company as the date of the firm's stance, since this is likely to reflect when the firm has taken action; there is sometimes a slight delay between a company's press release and their social media posting. In some instances the firm has not released a statement, or in the case that the firm has not dated their press release, I use google news search to find the earliest mention of the company's actions that correspond to the press release. I follow up with secondary time variables on the date in which the

company makes additional changes to its position.

This means I made several changes to the coding of some companies from the original Sonnenfeld database using this rubric, since the Sonnenfeld database updates all company status points as soon as the company has moved to a new position; i.e. Heineken would be listed as 'withdrew' even though the company only moved to this status in August 2023. To exemplify; if I found an earlier press release from the company stating that the company would be waiting to see how the situation progresses, I would code the company as 'D: Buying time' as its first action, and subsequent press releases/social media releases I would then code the company as 'C: Scaling back' if the company indicated it was taking further action. For example, in the case of the Polish company 'CANPACK SA', the company issued a statement on the 25th of February 2022 that represented the companies attempt to defy demands for exit or reduce activities (F: Digging in) which represented its first action, however on the 11th of April 2022, the company issued a follow-up statement that categorized the company as 'D: Buying Time as its second action. I followed each company's actions over the course of 2022, and in some instances there are several time points of action along the spectrum, with some companies making several moves in 'reverse' and others making several moves towards 'withdrawal'. For the purposes of this paper, I only consider two significant time points as the changes in position to grade company withdrawal. To construct the time variable for the network analysis I used 5-day increments due to computational limitations.[20]

[20] In the appendix I use a proxy for social pressure and social awareness, here I follow Lu, Huang, and S. Li 2022 and proxy for this using internet freedom and internet subscription rates (`intfreedom, internetsubscription`) from Freedom House. This measure for reputation in particular is backed up by Dyck, Volchkova, and Zingales 2008. See the appendix for more on this.

### 5.2 Reputation: Social Media Statements

I include a variable `socialmediastatement` if the company issued a social media statement in relation to a) the conflict, and b) the companies response. This represents a direct and consumer-facing statement of response to the public. I searched social media platforms for statements (the main platforms: X, Facebook, Linkedin and Instagram) and the variable is coded 1 if the company made a press release in relation to the conflict and the company's actions on any of these platforms, and 0 otherwise. As some companies issued statements in different languages, or used graphics/pictures of statements, and this would be missed by any search with key-words or scraping attempting, I manually searched each company handle for all of 2022 and 2023 statements. Importantly, the company must issue the statement from their main account handle - several companies operate with multiple account handles (either in different countries or different subsidiaries of the company). The company is only coded '1' if they have used their main account, with the highest average followers, to issue their statement. If the company has posted in relation to the conflict but has not mentioned their companies stance on the conflict, or the actions the company is taking (i.e. the post only mentions humanitarian donations to Ukraine) the company is still coded as 0. In some other instances the firm may be posting about their analysis of the effects of the crisis on GVCs or markets as research/business information - this is especially true of major consulting and legal firms. These posts do not indicate the firm's position on the conflict, and therefore these firms will also be coded 0. This methodological issue reinforces the need for manual coding since attempts to code based on mentions of Ukraine or Russia would ignore the context of these terms used by the company. For the reputation channel to be tested, the firm must make reference to its position on the conflict.

I coded the company as 1 if it had made a press release (`firmstatement`) in relation to the conflict and company's actions, but these needed to be explicit press releases, blog posts or featured news on the company's website. Quarterly reports to shareholders were not included.[21] One issue worth mentioning is that some companies have subsequently deleted their statements on the conflict from their websites; I have not updated the dataset to include this information, but it is certainly meaningful and an interesting avenue for future research.[22]

I follow current practices in the literature and include the following control variables in my main model to ensure that the effect of reputation is not confounded by the effects of other variables that could substantially affect firm behaviour, or the capacity of firms to withdraw from Russia: `country of HQ, industry, sub sector, firm size`. The exact market share of firms in Russia is difficult to find and only some firms disclose their exposure in Russia through statements or financial reports. In line with Pajuste and Toniolo 2022, I used the KSE variable and supplemented with manually coded geographic revenue exposure (`GeoRev` to proxy for firm exposure to Russia.[23] For the main model, I control for firm size – which is the natural logarithm of the market value of equity (Computstat items: `prccfx csho`.[24] In the appendix I include additional controls, such as whether or not the company's headquarters is located in a country that is allied to Russia (Correlates of War alliance data), as I am al-

[21] In financial reports, companies would often refer to the conflict to reassure stakeholders of the limited impact of the Russian invasion on profits.

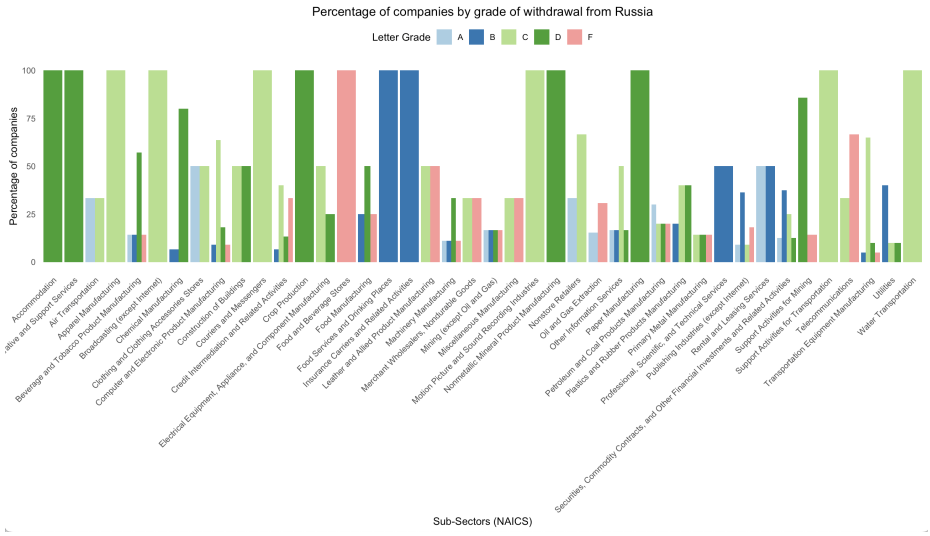
[22] The deletion of statements from company websites is a big issue for reproduceability, the existence of archival news articles tracking these statements is the only way to now confirm they were issued - although often these news articles are not accurately dated for the precise date the company issued a statement.

[23] See appendix for more information on how this variable was coded.

[24] The natural log of the market value f equity is an established way to operationalize firm size (Lins, Servaes, and Tamayo 2017, I use it because using other measures of firm size (e.g., the natural logarithm of the number of employees) as control variables leads to considerable multicollinearity, and this information is not consistently provided by companies in the dataset.

ready employing country-level fixed effects, this has no impact on the statistical significance or substantive effect of industry rate on withdrawals.[25]

I use fuzzy matching (`fedmatch` and `fuzzyjoin`) to pair companies in datasets using the Jaro-Winkler algorithm which takes into account the length of strings, the number of shared letters, the number of transpositions of shared letters (e.g. example and example would be 1 transposition), and the number of matching letters among the first four letters. Jaro-Winkler is reliable in cases when the imperfections happen mainly at the end of the string, for example through additional words (Company ABC International vs Company ABC Int). I double checked all entries, and in the case of poor matches, I re-coded entries for better matches. Below are some descriptive statistics, including the top 15 firms by revenue exposure to Russia, and the distribution of firms across sectors in my sample:




[25] Although on this note, future research could investigate whether the public in allied states influences firm behaviour in countries subject to sanctions.

**Top 15 Firms by Revenue Exposure to Russia**

| Company Name | Revenue Exposure to Russia (%) | Country of HQ | Status | Industry |
|---|---|---|---|---|
| petro welt technologies | 95.63 | Austria | exited | Energy, oil and gas |
| spar | 63.74 | Netherlands | leave | FMCG |
| nokian tyres | 62.85 | Finland | exited | Automotive |
| hyve group plc | 60.59 | United Kingdom | exited | Events, Education |
| kamat | 51.17 | Germany | stay | Engineering |
| veon | 51.04 | Netherlands | leave | Telecom |
| voith | 41.23 | Germany | stay | Engineering |
| eltek | 38.44 | Norway | leave | Electronics |
| krka | 37.16 | Slovenia | stay | Pharma, Healthcare |
| global fashion group | 35.41 | Singapore | exited | Fashion and leisure |
| ht troplast | 32.77 | Germany | stay | Manufacturing |
| maire tecnimont | 32.71 | Italy | wait | Engineering |
| aps energia | 30.67 | Poland | leave | Energy, oil and gas |
| frigoglass | 30.59 | Greece | stay | Electronics |
| orion pharma | 27.27 | Finland | leave | Pharma, Healthcare |

## 6 Research Design

To test the effect of reputation on firm behaviour, I use four different modelling strategies to estimate the effect of sector diffusion withdrawal rates on the speed of firm statements to withdraw. I run a series of two-way fixed effects (2FE) regression models on my full sample of firms over a period of six-months to control for unobserved heterogeneity in the sector and country-level, since both of these variables exert independent effects on firm decisions to withdraw. Second, I re-run these models using the sub-sector instead of sector-level to test the effect at a more granular level of industry-type. I also split my sample into early-movers and late-movers, since my argument has a strong temporal dimension; reputa-

tion is a positional good where the firm that chooses to withdraw early looks better, but makes their nearest-neighbours look worse. At the first stage this should encourage firms to withdraw quickly.

The quest for social prestige or the enhancement of self-image is, in itself, a zero-sum game. In the parlance of sociologists it is a positional good. The buyer of a hybrid car feels and looks better, but makes his neighbours (both buyers and non-buyers of hybrid cars) feel and look worse. This is a 'reputation stealing' externality. In the limit, when everyone behaves in a socially responsible way, no one gets credit for it. Therefore in the second stage I test the effects of firm withdrawals on late-movers. In the appendix I check the model's sensitivity by using a scale of withdrawal along the dimensions of 'withdraw', 'scaling back', 'suspension', 'buying time', and 'digging in'. I check for robustness using 2FE logit models.

### 6.1 Intra-industry diffusion effects of Reputation

Given the different structures of the data, I utilize a few different types of estimation models. First, when withdrawal is a binary indicator, I estimate logit models. Second, when I treat withdrawal as a scale of continuum between 'digging in' and 'withdrawing' I use OLS regression. Both models include fixed effects at the country level. I include robust standard errors clustered at the industry level. I use a Cox proportional hazard model due to the time-dependent nature of the outcome variable. Specifically, I estimate semi-parametric Cox proportional hazard models, the preferred method of duration analysis (Therneau 1997).

The cox PH model is suitable for estimating models with grouped event history whereby the observations of firm statements are temporally dependent Beck, J. N. Katz, and Tucker 1998. I report robust standard errors adjusted for

clustering at the country level and industry level. I include coefficient estimates in Table X1. The cox PH model translation is intuitive; a positive coefficient in the table implies a quicker rate of withdrawal by days (a higher hazard rate) and a negative coefficient implies a slower rate of withdrawal by days (i.e., lower hazard rate). I re-coded the 0 variable which implies that the firms withdrew on the first day to be 0.5 which indicates the firm left within the first day[26]

The unit of interest here is firms. The independent variable is the diffusion effect of firms within an industry issuing a statement on the firm's position. I construct a variable `industryrate` that captures the behaviour of other companies within the same industry. This time-varying covariate represents the number of other companies within the same industry that have issued statements before the focal company at each time point. For each time point (day) in the dataset, I generated this 'intra-industry diffusion effect' measurement.

My models of withdrawal duration following Russia' invasion on February 24th 2022 is as follows:

$$h(t|X) = h_0(t) \exp\left(\beta_1 \times \text{diffusion_rate} + \beta_3 \times \text{variable}_3 + \beta_4 \times \text{variable}_4 + \beta_5 \times \text{variable}_5\right) \tag{1}$$

[26] More granular analysis could be achieved by using hour-to-hour withdrawals since some firms moved within a 12-hour window of their competitors, although this could only be collected for some firms in the sample.

$$h(t|X, \text{country-year}) = h_0(t, \text{country-year}) \exp(\beta_1 \times \text{diffusion_rate} + \beta_3 \times \text{variable}_3 + \beta_4 \times \text{variable}_4 + \beta_5 \times \text{variable}_5) \quad (2)$$

The hazard function $h(t)$ at time $t$ is the instantaneous rate at which events occur.

$$\texttt{Surv}\ (\textit{time}, \textit{issuedstatement}) = \texttt{industryrate} + \texttt{firm size} + \texttt{Georev} + \texttt{NAICS(subsector)} + \texttt{factor(country)}) \quad (3)$$

Where *issued statement* represents the vector of independent variables at continuously measured time points. The cox PH model is suitable for this type of analysis since it allows me to estimate the parameters of the intra-industry diffusion effect of movement in the presence of unknown or complicated time varying hazard. Data is indexed by day following withdrawal, whereby the discrete hazard on day 1 for firm $i$ is the probability the firm will issue a statement on withdrawing that day.

The key assumption underpinning the Cox hazard model is proportional hazard; a change in the explanatory variable (that is, the industry rate of exit),

has an effect on the hazard that is both proportional and constant over time. While this is an important assumption that is well-suited to certain tests and types of relationships between variables, it is an assumption that does not always hold in politics. The explanatory variable in my argument waxes and wanes over time, where reputation is a positional good and the benefits of exiting diminish as the pool of leaving firms grows in proportion to remaining firms, and time remained in Russia goes on - such that late exiting firms will not be rewarded in the same way that early exiting firms were. A late exit will only remind audiences that the firm remained for a significantly longer time than their peers, and their motives are unlikely to be aligned with humanitarian or altruistic reasons. This may alter the conditions under which exiting occurs. This conjecture suggests that the industry rate should be associated with larger coefficients for early withdrawals, but smaller coefficients for late withdrawals. I therefore account for this assumption violation issue and time-varying effects of industry rate by running relevant specification checks based on the scaled Schoenfeld residuals for the models estimated. This information is included in the appendix, alongside tests for the Variance Inflation Factor (VIF) to assess how much the variance of the estimated regression coefficient increases due to collinearity. The results of these robustness and assumption violation tests can be found in the appendix.

## 7 Empirical Analysis

### 7.1 First Action Withdrawals

Table 1 shows the results of the PH model with country fixed-effects with a random inclusion of a sample of industry and country effects due to space constraints.[27] The variable `industryrate` represents the diffusion effect of firms within the industry moving to withdraw in their first action. The coefficient of

[27] Full table available in Appenix A

3.192 means that for a one-unit increase in the `industryrate`, the hazard of issuing a withdrawal statement increases by a factor of approximately 24.34 times. The hazard ratio is the exponentiation of the coefficient and indicates the multiplicative change in the hazard for a unit change in the predictor. The hazard ratio for model 2 is approximately 308.59. This significantly higher ratio suggests that in the context of Model 2, the likelihood of a firm withdrawing from Russia increases dramatically – to about 308 times higher – for each additional industry peer issuing a statement. In simpler terms, this suggests that an increase in the diffusion effect of firms within the industry moving to withdraw is associated with a significantly higher risk of a company issuing a withdrawal statement following the Russian invasion. The findings here indicate strong evidence for H1a. The negative coefficients for UNGC membership and ESG scores may indicate that firms with higher commitments to global standards and sustainability are more cautious or take longer to withdraw, possibly due to more complex stakeholder engagements or reputational considerations. Firm size appears to have a less clear impact on the decision to leave the Russian market, with different directions of influence based on company scale. This might indicate that other factors, such as operational complexity or market engagement level, play a role in larger firms.

Table 2 adds in additional control variables from the KSE database and data on CMC from Refinitiv Eikon, and considers sub-sector diffusion effects as coded through NAICS codes. In this new table, we see the results from a Cox Proportional Hazard Model that explores the 'Sub-Sector Diffusion Effect' on firm withdrawal from Russia. Model 1 is a bi-variate regression, model 2 adds in controls but considers only country-level fixed-effects, and model 3 adds in controls but considers both country and industry level fixed effects. I calculate the

**Table 1.** Cox Proportional Hazard Models: First Action Withdrawals

| | Survival Risk | |
|---|---|---|
| | Withdrawal | |
| | (1) | (2) |
| Sector Diffusion Effect | 3.034*** | 5.732*** |
| | (0.242) | (1.666) |
| Firm Size: Mid CMC | | 0.194 |
| | | (0.347) |
| Firm Size: Smal CMC | | −0.529 |
| | | (0.545) |
| UNGC member | | −0.171 |
| | | (0.303) |
| Allied State HQ | | |
| | | (0.000) |
| ESG score | | −0.017* |
| | | (0.009) |
| FE (Country, Industry) | No | Yes |
| Observations | 1,275 | 947 |
| $R^2$ | 0.081 | 0.020 |
| Max. Possible $R^2$ | 0.880 | 0.396 |
| Log Likelihood | −1,295.528 | −229.336 |
| Wald Test | 157.630*** (df = 1) | 17.990*** (df = 5) |
| LR Test | 107.092*** (df = 1) | 18.838*** (df = 5) |
| Score (Logrank) Test | 185.775*** (df = 1) | 18.976*** (df = 5) |
| *Note:* | | *p<0.1; **p<0.05; ***p<0.01 |

hazard ratios for the sub-sector diffusion effect in each model: Model 1: HR = exp(2.227), Model 2: HR = exp(2.968), Model 3: HR = exp(1.758). The hazard ratio for Model 3 is approximately 5.8. This suggests that for each additional company within the same sub-sector issuing a statement, the likelihood of a firm withdrawing from Russia is about 5.8 times higher, compared to when no additional firms have issued statements. The Sub-Sector Diffusion Effect consistently shows a strong positive impact on the likelihood of firm withdrawal from the Russian market across all three models. This indicates that firms are significantly influenced by the actions of other firms within their sub-sector. Predictably across table 1 and table 2, firms from China, India and Turkey, are much less likely to be influenced by sub-sector diffusion effects.

**Table 2.** Cox Proportional Hazard Models: Sub-Sector Effect on Withdrawal with Added Controls

| | *Dependent variable:* time_diff | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Sub-Sector Diffusion Effect | 2.227*** | 2.968*** | 1.758*** |
| | (0.074) | (0.204) | (0.108) |
| Firm Size Mid | | 0.417 | 0.195*** |
| | | (0.395) | (0.064) |
| Firm Size Large | | | |
| | | (0.000) | (0.000) |
| Number of local companies in Russia | | −0.050* | −0.022 |
| | | (0.029) | (0.030) |
| Revenue Exposure in Russia (2021) | | 0.002*** | 0.002*** |
| | | (0.0003) | (0.0002) |
| Revenue Exposure in Russia (2022) | | −0.003*** | −0.004*** |
| | | (0.0004) | (0.0004) |
| Allied State HQ | | −17.368 | −16.716 |
| | | (1,651.270) | (524.507) |
| FE (Country, Sub-Sector) | No | Yes | Yes |
| Observations | 6,943 | 3,051 | 3,051 |
| $R^2$ | 0.132 | 0.393 | 0.241 |
| Log Likelihood | −34,367.220 | −13,450.440 | −13,791.580 |
| Wald Test | 899.850*** (df = 1) | 730.820*** (df = 66) | 267.780*** (df = 6) |
| LR Test | 980.879*** (df = 1) | 1,525.167*** (df = 66) | 842.889*** (df = 6) |
| Score (Logrank) Test | 900.512*** (df = 1) | 1,332.637*** (df = 66) | 640.870*** (df = 6) |

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

Figure 1 illustrates the effect of within-sector firms withdrawing using a Kaplan-Meier survival curve. In assessing the baseline hazard of a firm withdrawing following within-sector withdrawals, this figure shows that many firms withdrew quickly within the first 20 days, and there is some flattening in certain periods following a continuation of withdrawals. The median survival of firms in Russia in my sample of companies following within-sector firm withdrawals is 77 days, with the hazard of withdrawals stabilizing after the 70 day mark for 'first action withdrawals'. Figure 2 desegregates this model into a Kaplan-Meier plot with industry groups. Real-estate groups had the lowest median survival rate at just six days, followed by Communication services at 18 days. In contrast, Information Technology had the longest at 77 days, and followed by Industrials at 66 days.

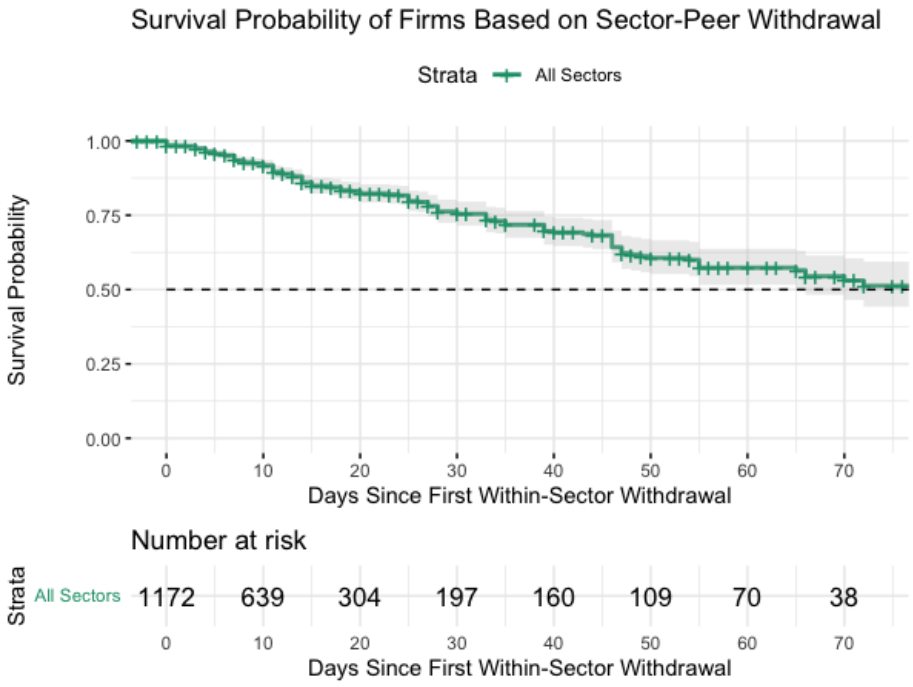



**Fig. 1.** Kaplan-Meier Plot: Effect of Within-Sector Withdrawals on Time of Withdrawal of Firms: First 75 Days

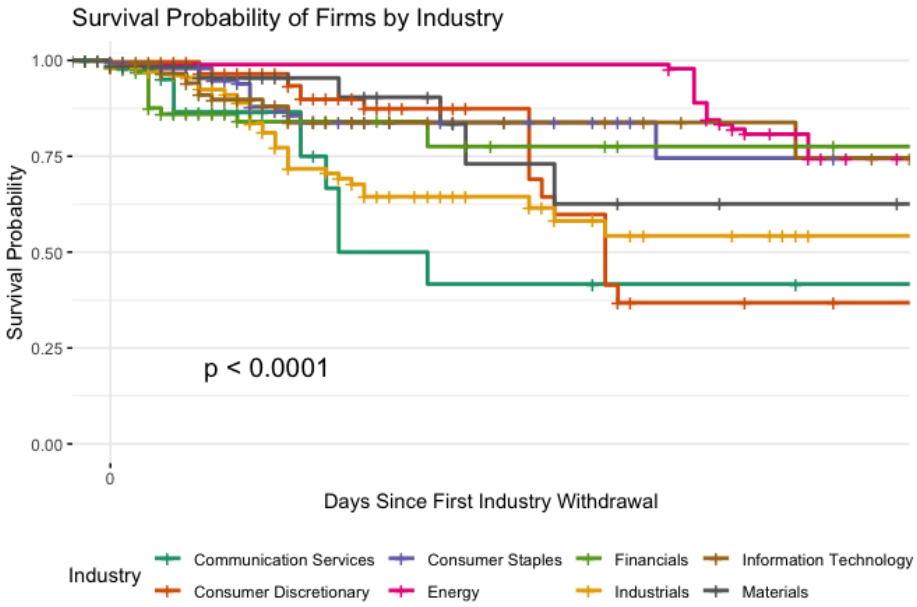



**Fig. 2.** Kaplan-Meier Plot: Sector Effects (First 70 Days)

### 7.2 Late Withdrawals

In investigating the diffusion effect of sector withdrawals on company behaviour, I tested the effect on late withdrawals, or companies who chose to withdraw as a second action. I specifically collected this data to investigate how more nuanced sectoral dynamics influenced individual firm decisions over time, and if it is the case that reputation plays a pivotal role in the early stages of the invasion when public scrutiny is intense, and the social capital rewards/avoidance of social sanctions is high, then I would expect to see a positive influence of this indicator in the early stages. For second actions, when firms have already signalled their willingness to stay, when the population of firms that demonstrated early socially responsible behaviour has increased, there is no credit for withdrawal. Withdrawal might help mitigate public backlash on the firm eventually, however

there is always the risk that signals of withdrawal at later stages may remind consumers how long it took the company to withdraw. Therefore, I would expect that for late withdrawers (or those who withdrew as a second action), the socially optimal incentive rate of withdrawal should decrease or lose significance, and the diffusion effect should wane.

Table 3 shows the results of this test, with model 1, a baseline model, and Model 2 incorporating additional controls and fixed effects for country and industry to account for unobserved heterogeneity that could influence late withdrawal decisions. The findings do not necessarily confirm `hypothesis 1b` that the diffusion effect would weaken over time, but rather suggest that the effect of peer-withdrawals discourages further within-industry withdrawals. This is a different outcome than I hypothesised. In Model 1, the 'Sector Diffusion Effect' was negatively associated with the likelihood of a firm's late withdrawal (-0.047). This means there's a 4.6% decrease in the hazard for each additional firm withdrawing in the same sector. In Model 2, this effect is larger (0.371), suggesting a 31% decrease in the hazard of a firm withdrawing in second-stage actions. The logic of a reputation stealing externality holds here, but is stronger than I anticipated, where the reputational impact of not following the initial wave of withdrawals might lead to greater hesitancy. The alternative explanation is that remaining firms may view the trade-offs of remaining more favourably as the exodus of competitors leaves a profitable vacuum for expansion. This is certainly the case for China, which has cultivated an expanding space as the substitute supplier of goods and service in Russia. Chinese state-owned enterprise, Shanghai Automotive Industry Corporation, has expanded rapidly in the manufacturing industry in Russia.

**Table 3.** Cox Proportional Hazard Models: Second Action Withdrawals

| | Survival Risk | |
|---|---|---|
| | Withdrawal | |
| | (1) | (2) |
| Sector Diffusion Effect | $-0.047^{***}$ | $-0.371^{***}$ |
| | (0.005) | (0.041) |
| UNGC member | | −0.133 |
| | | (0.311) |
| Big Firm Dummy | | −0.328 |
| | | (0.333) |
| ESG score | | −0.006 |
| | | (0.010) |
| Allied State HQ | | |
| | | (0.000) |
| FE (Country, Industry) | No | Yes |
| Observations | 1,267 | 944 |
| $R^2$ | 0.092 | 0.190 |
| Max. Possible $R^2$ | 0.953 | 0.489 |
| Log Likelihood | −1,880.687 | −218.069 |
| Wald Test | $90.420^{***}$ (df = 1) | $87.040^{***}$ (df = 4) |
| LR Test | $121.628^{***}$ (df = 1) | $198.521^{***}$ (df = 4) |
| Score (Logrank) Test | $98.544^{***}$ (df = 1) | $90.196^{***}$ (df = 4) |
| *Note:* | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

### 7.3 Social Media Statements

Next I analyze the effect of firm exit decisions on issuing social media statements using probit models. The dependent variable here is a dummy variable indicating whether the firm issued a social media statement referencing their position or actions in response to the war. I use three versions of the explanatory variable; i) total company withdrawals, which represents the total number of companies who issued a statement to withdraw in either their first or secodn action, ii) early withdrawals, which represents companies that issued statements to withdraw early as a first action, and iii) late withdrawals, which represents companies that issued statements to withdraw as second action. I control for ESG, firm size (company market cap), UNGC membership, and Russian-allied state HQ with two-way industry and country fixed effects. Some observations have been dropped because of the restricted sample of hand-coded social media observations. The results in Table 4 for the fixed-effect models with controls shows that withdrawing companies are less likely to issue social media statements than remaining statements.

To calculate the marginal effect of a probit model we assume that it is roughly equal to 20% of the coefficient value. This is an approximation and assumes an average value for the probability density function. In Model 2, there is a highly significant and negative effect of withdrawing firms on the probability of issuing a social media statement. Total withdrawals decrease the probability of issuing a social media statement by about 1.98%. Similarly, withdrawing firms are -3.56% less likely to issue a social media statement than remaining firms, and the decision to withdraw late has a slightly less significant and substantive effect on issuing a social media statement, at -1.94%. These results imply the opposite of my `hypothesis 1c`, remaining firms are much *more* likely to issue social

media statements to protect their reputations than withdrawing firms. Indeed, the finding that this is especially true for immediate withdrawals where public scrutiny was highest than late withdrawals, suggests the logic of reputation stealing externalities does hold, but instead of motivating withdrawing firms to 'signal boost' to their audience and reap reputational rewards, remaining firms are more likely to capitalize on this strategy to prevent reputational losses.

The finding suggests these companies employ this strategy as a way to deflect or obfuscate their own position on Russia without requiring them to take the costly action of withdrawing. In terms of my control variables, membership in the United Nations Global Compact seems to have a small positive effect on issuing a statement ( 1.14%), and the environmental, social, and governance score has a small positive effect on issuing a statement in Models 2 and 3, but not model 4 (0.04%). Together, this suggests that firms with high ESG scores and membership to voluntary regulatory bodies are more performative than their peers. As would be expected, firms headquartered in states allied to Russia are significantly less likely to issue a social media statement. In the appendix I examine the marginal effects of firm size on likelihood of issuing a withdrawal statement; when controlling for relevant covariates, firm size has little impact on firm decisions to issue a statement to withdraw.

**Table 4.** Probit Models: Effect of Firm Exit Decisions on Issuing Social Media Statements

| | Issues Social Media Statement | | | |
|---|---|---|---|---|
| | Statement Dummy | | | |
| | (1) | (2) | (3) | (4) |
| Total Withdrawals | −0.001 | −0.099*** | | |
| | (0.031) | (0.036) | | |
| Early Withdrawal | | | −0.178*** | |
| | | | (0.050) | |
| Late Withdrawal | | | | −0.097* |
| | | | | (0.054) |
| UNGC member | | 0.056* | 0.057* | 0.046 |
| | | (0.034) | (0.034) | (0.059) |
| Big Firm Dummy | | 0.067 | 0.067 | −0.004 |
| | | (0.041) | (0.041) | (0.070) |
| ESG score | | 0.002** | 0.002** | 0.001 |
| | | (0.001) | (0.001) | (0.002) |
| Allied State HQ | | −0.388 | −0.420 | −1.030*** |
| | | (0.280) | (0.280) | (0.258) |
| Constant | 0.496*** | 0.247 | 0.296 | 1.256*** |
| | (0.016) | (0.263) | (0.263) | (0.274) |
| FE (Country, Industry) | No | Yes | Yes | Yes |
| Observations | 1,280 | 950 | 951 | 324 |
| Log Likelihood | −929.956 | −531.843 | −529.365 | −133.578 |
| Akaike Inf. Crit. | 1,863.911 | 1,165.687 | 1,160.729 | 343.155 |

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

### 7.4 Shareholder returns: Rewarding Withdrawals

The argument I have made in this paper is that there is a strategic interaction occurring between firms who must adjust their stance on Russia in response to their competitors - not because the withdrawal of a competitor signals information about the economic risks of remaining in Russia - but because the withdrawal of a competitor increases the reputational costs of remaining. The competing explanation to reputation as a motivator is that evidence of strategic withdrawals driven by peer and competing firm withdrawals is motivated by a learning strategy through which the manager of the firm treats competing firm withdrawals as a source of information about the relative business risks of remaining and/or opportunities from withdrawing. This would be an informational cascade effect where a firm, having observed the actions of their competitor, chooses to follow their behaviour, regardless of the firm's private information, because they believe their competitor may be more knowledgeable (Bikhchandani, Hirshleifer, and Welch 1992). This 'imitation' act can reap rewards for businesses in uncertain markets and business environments (Cyert, March, et al. 1963; DiMaggio and Powell 1983).

In order to probe the mechanism for company withdrawals; whether firm action is motivated by information or reputation,[28] I consider three possible sources of information that would suggest firms are purely motivated by economic incentives and peer learning of the possible rewards/losses of leaving/remaining. The first is immediate investor reactions to firm statements to withdraw. Do investors reward or punish firms for socially responsible behaviour? Since evidence of consumer behaviour and/or reactions may be delayed, the best source of information for competing firms. There is already robust evidence that this

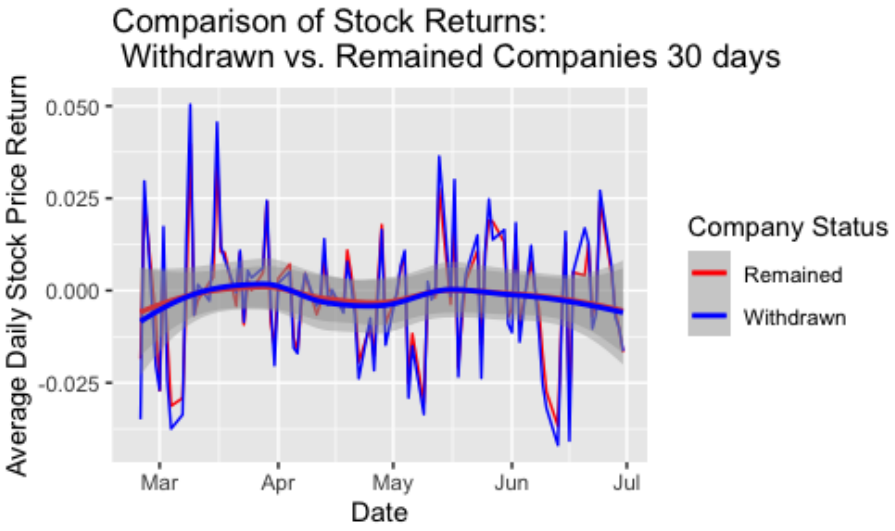
[28] To reiterate, this paper is concerned with statements to withdraw or remain, and not confirming true exit. For identifying true exit see Wellhausen and Zhu (2024).

is not the case. Lu, Huang, and S. Li (2022) investigate stock price returns of firms that announce withdrawals over a 1-3 day period post withdrawal. If withdrawing Russian firms are rewarded by investors who place a premium on the socially-responsible behaviour of the firm, or alternatively these shareholders regard ESG manoeuvres as both neutral and cost-less, then stock prices should increase. This information would then be a helpful signal for competing firms to also withdraw. This test is confined to a short-period of firm withdrawals given the decision to withdraw amongst firms occurred within a short-period of time. It is possible, and indeed true, that some firms waited much longer to withdraw. I compute the average change in the percentage return of stock prices using the daily price return.[29] This data is collected in USD dollar returns. Please see appendix for these results.

I collected stock price information for publicly-listed companies from one day prior to the initial date of invasion (23nd of February) until July 2023, amounting to just over 112,000 observations over this period. I collected data on daily stock price, stock price high of the day, stock price low of the day, flow of daily prices, share turnover, and daily percentage change in stock price compared to the previous day. I am primarily interested in the daily percentage change in stock price between companies that remained and companies that withdrew, and the share turnover, as representative of short-term trading of the company's stocks. Initial T-tests comparing treated companies (withdrawing) to control (remain) show a steeper decline in stock prices for withdrawing firms versus remaining firms. The results suggest that the effect is acute in a 1-3 day period. When expanding to compare how company stock prices are affected by withdrawals in

[29] Unlike Z. Li and Disalvo 2023 I do not use the "abnormal return using the raw return minus the beta of the firm multiplied by the market return of the country where the firm's equity is traded", due to data source issues.

a 30 day window, there is no affect of withdrawing on stock price comparative to remaining. This suggests no financial reward for firms that withdrew; and for within-industry or within-country behavior it indicates that there was no 'information cascade' whereby firms withdrew after learning they could potentially gain from withdrawing. It also suggests that withdrawing firms were not inadvertently sending a signal to remaining firms that they had hidden information that withdrawing would be economically efficient.




This backs-up prior findings by ibid. These authors teased out whether investors punished firms for withdrawing (therefore reputation becomes a better explanation for withdrawals), or if shareholders reward firms for withdrawing (and economic benefits/incentives are a better explanation for withdrawals). Using propensity score one-to-one-matching, firms that belong to the same country, industry, and of similar size (measured by company market cap), are then matched to control firms. Using the stacked event-by-event difference-in-differences (DiD) Lu, Huang, and S. Li 2022; Cengiz et al. 2019; Baker, Larcker, and C. C. Wang 2022. The authors find similar results; investors neither reward *or* pun-

ish withdrawals. While these authors find no evidence for investors rewarding withdrawals, and cites this as evidence that the reputation mechanism is *not* in play, this is because the authors are framing reputation as a way to leverage economic gain, rather than as a way to avoid social sanctions and consumer backlash. Evidence of either greater stock price decreases in a 1, 2, and 3 daily period for firms that withdrew compared to firms that remained, or evidence that stock prices do not increase relative to firms that remained, suggests there is no economic reward for withdrawals.

### 7.5 Teasing out the Causal Role of Reputation

The results from these tests on industry rate and competing explanations lends strong support for the causal mechanism of reputation pressure induced by sector withdrawals on the rate of firm withdrawals. However, without data on measuring public attitudes towards specific firms and resulting firm withdrawals, it is difficult to truly evaluate the extent to which concern for reputation was motivating the decision-making of company executives. One way to do so would be to interview or survey company executives; however the likelihood of receiving honest responses that companies acted out of concern for reputation rather than true altruism would be low.

The second method would be to track public backlash online through geographically coded data on X (Twitter). There are two problems with this. First, with the closing of the academic-twitter API in 2022, and introduction of paid subscription models, obtaining this data is expensive. Second, there are serious concerns with how vocal individuals are on twitter and obtaining meaningful geographically coded data. The demographics of users on X, and the noise of the vocal minority can be a major distortion for this type of data (Gayo-Avello

2012). This is not to say this data has no use, indeed opinions on X by the vocal minority can disturbingly and rapidly inspire the silent majority (Castioni et al. 2022). Future researchers with more resources could test this using a subset of highly visible firms and tracing tweets/retweets overtime with key search terms related to "Russia" and measure this against withdrawal rates.

In the absence of this data, I explore another source of information that would reliably measure the effect of reputation on company decision-making and how companies managed their reputations; company statements. It has long been established in the literature on impression management, that organizations selectively communicate behaviour or actions to stakeholders that appeal to normative expectations (Fiss and Zajac 2006; Westphal and Graebner 2010). In turn, companies rely on stakeholder approval for success and survival. By issuing statements to communicate the companies socially responsible behaviour, they can secure positive stakeholder impressions and/or scape-goating and justifying poor behaviour (Ginzel, Kramer, and Sutton 2004; Soule 2018). The primary goal of issuing a company statement is to improve corporate reputation in a domain-specific area (McDonnell and King 2013; Carlos and Lewis 2018; Werner 2015). Existing research suggests these statements on company's pro-social behaviour does improve individual ratings of company reputations, and even more than this, public confessions from companies can, in some domains, improve public ratings of company reputations as a result of perceived moral authenticity (L. Z. Li and Soule 2022). All company statements from firms, either withdrawing or remaining, are likely related to reputational management. Above I used the presence or absence of these statements to evaluate whether withdrawing firms or remaining firms where more or less likely to issue statements. Remaining firms are much more likely to issue these statements, and the effect suggests these

statements are a critical resource for reputational management of remaining firms. In the next section I use computational text analysis to evaluate how the statements differed substantively between firms that issued statements to remain versus those that issued statements to leave.

## 8 Firm Statements

The statistical analysis reveals some interesting patterns in terms of how firms respond to each other's movements, and how decreasing public scrutiny over Russia has led to reversal in position of foreign firms along the withdrawal spectrum. Simply put, as consumer audiences pay less attention to the war overtime, some firms reverse their statement of withdrawal and entrench themselves further into Russia.

What's interesting is that some firms clearly allowed themselves room to move along the withdrawal spectrum; meaning they would be unlikely to suffer great penalties by returning to Russia, while others 'tied their hands' through their own public statements. Firms that anticipated returning to the Russian market, made three choices. The first was to avoid issuing a statement directly to the public through social media tools, preferring instead to issue a website statement, or make passing reference to the invasion and the company's response in a letter to investors or relevant stakeholders. The second was to avoid mentioning Russia in their statement at all - for example "the current situation" or "recent events" as opposed to using weighted language on the international legality of Russian actions in Ukraine, i.e. "Russia's invasion". The third was to reference compliance with laws and imposed sanctions to signify that the firm was not acting out of its own agency or free-will to leave Russia, but was instead following enforced regulations imposed by home governments. This stands in

contrast to the choice of many withdrawing firms to omit mention of compliance with sanctions, and instead suggest the move was motivated by the firm's moral compass and personal beliefs.

An key limitation to the above analysis of firm withdrawals, is that it's not entirely clear what is motivating firms to leave Russia. Are the firms leaving purely because they're complying with sanctions, potential market risks in Russia, or for moral reasons? Are they tying their hands when they do leave - by putting forward a firm and difficult to reverse position? Or are they using 'cheap talk' by avoiding accusatory and defamatory language against Russia? Indeed, identifying firms' preferences directly is a challenge given that firms often do not disclose information about these types of business operations. To overcome these methodological barriers, I investigate the content of firm statements on withdrawal from Russia that were issued by the company at the time of announcing action on Russia. Statistical models are hamstrung in their ability to show behavioural patterns. In this section, I employ computational text analysis of a sample of over 300 company statements on their initial decision to withdraw.

One way to identify differences in preferences when firms send hard withdrawal signals of 'tying their hands' and firms sending weak signals 'cheap talk' is to compare statements that were made by firms that withdrew completely in the initial stages versus firms that attempted to dig in or buy time. This type of analysis could reveal whether firms that choose to leave are systematically different from those who attempted to stay, or allow themselves room to return to the Russian market.

I constructed a corpus from manually scraped explicit company statements issued by the firm on the Russian-Ukrainian war between 2022 and 2023. To pull explicit statements in reference to the war I first consulted the Business and Human Rights Resource Center (BHRRC), which is an NGO that functions as an informal dispute mechanism for CSR claims against companies related to their ESG activities. In response to allegations, the center approaches companies and asks for direct responses, with systematic follow-ups to non-compliant firms. The BHRRC then collects data on respondents and non-respondents. Similiar to the Sonnenfeld, Tian, et al. (2022) and Sonnenfeld, Mylovanov, et al. (2023) Yale list, the platforming of issues by the BHRRC increases the salience of the allegations against the firm, and ramps up reputational costs (Thrall 2021). The BHRRC invited 400 companies operating or investing in Ukraine and/or Russia to respond to questions related to the firms activities and human rights due diligence. Companies were given two weeks to respond, and just 29 percent of firms approached issued responses (115 companies). Of those, just 43 provided full statements about their practices. I gathered these statements manually since some companies submitted PDFs, links to different HTML-style websites for full statements, or linked word documents. I gathered a further 257 statements (over 300 statements in total) by randomly sampling from my dataset of both withdrawers and remainers. I manually collected and coded these statements in a corpus.[30] By collecting statements from both remainders and withdrawals, I am able to study the relationship between reputation and the nature of withdrawal. Below are descriptive statistics about the characteristics of the firms in this corpus. Importantly, some firms have issued multiple statements over time, meaning that the influence of reputation over time changes. I used the

[30] A key limitation of the data collected is that firms that remain in Russia are much less likely to issue public statements (as illustrated above), meaning that the sample of remaining firms from which I scraped statements might be systematically different from the firms that remained and chose not to issue a statement.

most comprehensive statement from each firm, but included relevant follow-up statements unless the firm substantially moves position (i.e. from digging in to withdrawal) since the corpus is composed of first action respondents (not second action respondents). This ensures an accurate mapping of firm positions over time. I matched the list of firms to my existing database with relevant variables on the Sonnenfeld, Tian, et al. (2022) grading system, country of headquarters, industry, sub-sector, and firm size.

I used a battery of different text-as-data techniques to analyze the effect of concern for reputation on the nature of how firms withdrew or remained in Russia. Below I outline the two key methods; (i) word embedding and (ii) topic modelling.

### 8.1 Word-Embedding

Reputational concerns affect the nature of firm decisions to issue statements on withdrawing or remaining, but concern for reputational costs also influences how these firms carry out this decision. For instance, remaining firms are attempting to balance two audience demands and protect conflicting reputations. Remaining firms need to protect their reputation amongst their international consumer base and their Russian audience; this may require the firm to present itself as the 'good guy' without making negative references to Russia. These firms achieve this by either deflecting and re-orientating attention away from the firm's continued presence in the region by issuing statements claiming the firm is donating to humanitarian international organizations and NGOs, or by justifying their continued presence by making claims that the firm has a "moral responsibility" to their employees in Russia, or that (in the case of many pharmaceutical firms), it would be amoral to punish Russian citizens by withholding

essential medical supplies. On the other hand, the firm's continued functioning in politically-contentious regions in the future may require the company to project a 'political neutral' image to its other audience (Russian regulators, the Kremlin, or even interested Chinese parties). The firm achieves this in a few ways – the firm issues statements outlining that the firm's suspension of operations as being the direct result of a legal necessity to comply with sanctions, rather than arising out of the firm's own violation, or the firm's moral judgement on the situation. The firm avoids any politically-charged language in statements on their position, in some cases this means an avoidance of references to Russia entirely, where the statement is exclusively focused on humanitarian relief in Ukraine and compliance with "relevant sanctions". In more explicit ways, this means the firm avoids referencing the word "invasion" and in some cases makes no mention of "war" (Simon n.d.). [31]

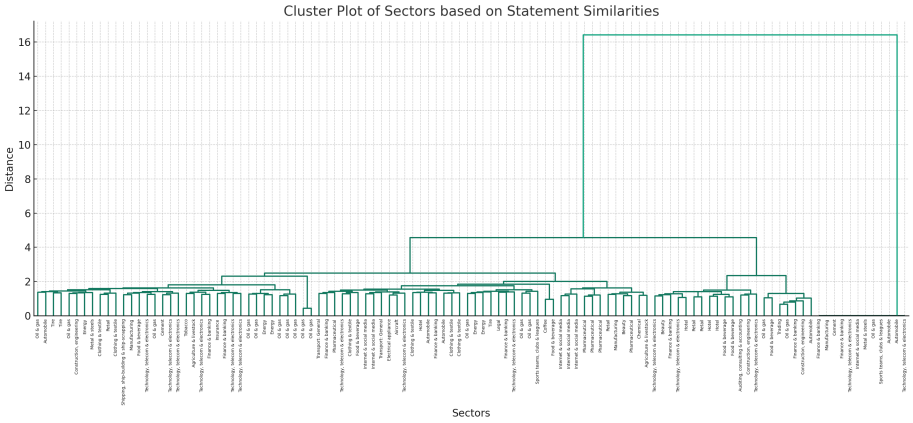
At the heart of word embedding is the idea that you can "know a word by the company it keeps" (Firth 1957, p. 11). Essentially, word-embedding offers semantic meanings by training a word embedding layer by converting a corpus into a term co-occurrence matrix. This matrix exploits "neural network architectures to learn the pattern of co-occurrences and derive a distributional representation of each word in the corpus in vector space" (Solis and Waggoner 2021). The two key modelling methods are GloVe (Pennington, Socher, and Manning 2014), and Word2Vec (Mikolov et al. 2013). Word-embedding is "generally robust to such choices for political corpora of various sizes and in various languages"(Rodriguez and Spirling 2022), alleviating concerns with a smaller corpus. There are few computationally expensive tasks involved in word-embedding, and as I am not

[31] This follows Kremlin directives framing the situation as a "special military operation on the territory of Ukraine" and Putin signing a law in March 2022 that criminalized reporting that opposed the war against Ukraine, with critical labels like 'invasion' and 'war' undermined the national interest, and constituted 'false information'.

interested in changes over time, I avoid the 'alignment problem' where word-embedding over temporal units diffuse the identification of a core vector space (Rodriguez, Spirling, and Stewart 2023; Hamilton, Leskovec, and Jurafsky 2016). To estimate word embedding I utilize the unsupervised GloVe method.

## 9 Intra-industry Reputation Effect: Cosine Similarity and Statement Mimicking

One way to verify the findings above that firms respond to their intra-sub-sector peers to preserve or protect their reputation is to evaluate whether the firms that do issue statements on their position in Russia mimic the language of their peers - i.e. copy-and-paste statements. To do this, I generate cosine differences between statement text of competing peers within an sub-sector. Specifically, I used the cosine as a similarity distance metric between statement vectors (i.e., firm statements converted into term-frequency vectors). For interpreting cosine similarity, measurements closer to 1 indicate higher degrees of similarity between statements, while numbers closer to 0 indicate lower degrees of similarity. The results are depicted below; there is a high degree of within-industry statement similarity for firms, although this differs between sectors.


Cluster Plot of Sectors based on Statement Similarities
Distance
0
2
4
6
8
10
12
14
16
Sectors

The dendrogram above illustrates the cosine similarities of company statements between sub-sectors. This hierarchical clustering analysis showing how closely sub-sectors are related based on their statement similarities. Each leaf represents an industry. The vertical lines connect industries or clusters of industries, with the height of the line indicating the distance or dissimilarity between them. Shorter lines represent more similar industries. Clusters are formed at different levels, shown by the height at which branches merge.

Using clustering analysis on the average cosine similarity of firms' statements within each sub-sector, I calculate the average by comparing each firm's statement with the statements of other firms in the same industry, excluding self-comparison. The highest average cosine similarity score, indicating that firms in this industry have very similar statements, are for Hotel and Construction. Internet and social media also have high average similarities, suggesting a commonality in the way firms in these industries communicate. Interestingly, the lowest average cosine similarity was for transportation, indicating more diversity in the statements within this industry. This potentially is due to the logistical difficulties of withdrawing given shipping and transportation infrastructure.

| Industry | Average Cosine Similarity |
|---|---|
| Agriculture & livestock | 0.327073 |
| Automobile | 0.289400 |
| Beauty | 0.404803 |
| Clothing & textile | 0.344414 |
| Construction, engineering | 0.447807 |
| Energy | 0.346475 |
| Finance & banking | 0.364563 |
| Food & beverage | 0.328595 |
| Hotel | 0.455049 |
| Internet & social media | 0.475232 |
| Manufacturing | 0.361096 |
| Oil & gas | 0.286921 |
| Pharmaceutical | 0.393245 |
| Retail | 0.402729 |
| Technology, telecom & electronics | 0.387348 |
| Tire | 0.321764 |
| Transport: General | 0.239023 |

The cosine-similarities are suggestive of a within-industry trend indicating that companies may mimic each others behaviour to protect their reputations.

## 10 Cosine similarity and criticism of Russia

In order to test whether firms that issued statements to remain and firms that issued statements to withdraw from Russia relied on different strategies in order to protect and manage their reputations amongst their relevant audiences, I use word-embedding techniques to analyze the content and differences between these

firm-types. I have argued that firms that remained in Russia sought to pander to both their consumer audiences and preserve their relationship with Russia in order to ease re-entry into the market and avoid risking future prospects. In order to do this, these firms will attempt to refocus consumers on their 'good' behaviour by emphasising their humanitarian contributions, leverage their relationship with UN-adjacent MSIs, and reassuring Russian audiences by using less politically-charged language, less criticism of Russia and more emphasis on legal obligations than moral initiative. I test for the latter by using sentiment analysis, term frequency and cosine similarity for embedded terms around 'Russia' and 'Ukraine'. I used TextBlob, which is a popular library for sentiment analysis. *TextBlob* determines sentiment based on polarity scores, which range from -1 (most negative) to 1 (most positive). Remainders had higher positive polarity sentiment scores for references to Russia, while Withdrawers scored lower on polarity sentiment.[32]

### 10.1 Computational Dictionaries: Linguistic differences

An economic explanation for why companies chose to either remain or leave would suggest that MNCs exist only to make money for shareholders through the provision of either goods or services. For companies, this would suggest they would avoid using any politically charged language, or engage in political activities when operating in Russia, and in their statements to leave. My argument suggests the opposite of this; that withdrawing firms would attempt to capitalize on their reputational gains and make stronger, moralistic statements with inflammatory language towards Russia, while remaining firms will attempt to divert international consumers attention away from their presence in Russia while maintaining cordial relations in Russia. I would therefore expect remaining firms

[32] See Appendix.

to be more likely to employ terms like 'compliance', and less likely to use terms such as 'invasion' or 'condemn'.

To test this, I extract the average word embeddings for each statement from withdrawing and remaining firms. I computed the cosine similarity between these average embeddings of the new terms: "invasion/invaded", "condemn", and "compliance, comply, complying, sanctions" using used principal component analysis (PCA). I then compare the average similarity scores for the two groups (Remained vs. Withdrew) for each set of terms.

| Firm | Invasion | Invaded | Condemn | Compliance | Comply | Complying | Sanctions |
|---|---|---|---|---|---|---|---|
| Remained | 0.813 | 0.908 | 0.904 | 0.679 | 0.734 | 0.794 | 0.602 |
| Withdrew | 0.843 | 0.930 | 0.921 | 0.651 | 0.716 | 0.778 | 0.580 |

Both groups have similar levels of association with terms like "compliance", "comply", "complying", and "sanctions", however the withdrawal group has a slightly higher association with the terms "invasion", "invaded", and "condemn" compared to the remainers group. The next piece of evidence involves the general sentiment of how these companies discuss Russia and Ukraine respectively. Once again, remaining firms are likely to employ much more positive and politically neutral language overall than withdrawing firms. Additionally, I would expect greater positive sentiment towards Russia from the remaining firms, and more positive sentiment to Ukraine from the withdrawing firms. I extracted the relevant sentences and computed the sentiment scores using TextBlob for each statement, where the sentiment score ranges from -1 (most negative) to 1 (most positive).

| Action | Avg Sim - Rus | Avg Sim - Rus' | Avg Sim - Ukr | Avg Sim - Ukr' | Avg Sent |
|---|---|---|---|---|---|
| Remained | 0.651 | 0.684 | 0.776 | 0.735 | 0.056 |
| Withdrew | 0.676 | 0.700 | 0.814 | 0.766 | 0.033 |

**Table 5.** Summary of Grouped Actions and their Metrics

Overall, remaining firms issue statements with much higher overall positive sentiment scores, and with slightly higher positive scores towards Russia than withdrawing firms. The results here are modest, and for more granular insight into the themes employed by firms in their statements I use a topic-modelling approach.

### 10.2 Topic Models: use of humanitarian language

I have argued that when a firm within a sector issues a statement to withdraw from Russia, they create reputational-stealing externalities on peers within their sector, motivating these firms to quickly signal their intention to exit to avoid consumer backlash. I also suggest that firms that remain do so because they are concerned with not only their international consumer base, but also their Russian and aligned-state audiences. Firms that remain would be subject to greater reputational costs from consumers, making their decision to remain very costly. If the reputational mechanism that causes this cascade effect is at play, then we should expect that remaining firms will attempt to divert attention away from their decision to remain, and keep both audiences (international consumers and Russian-allied actors) on side. It is crucial to the argument that reputation matters for all firms - not just withdrawing firms. It is not the case that firms that remained were insensitive to reputational costs; instead they used specific strategies to deflect attention and preserve their reputation. The cosine similarities above suggest remaining firms 'learnt' and subsequently adopted these strategies.

Using topic-modelling, I highlight how remaining companies and withdrawing companies differed fundamentally in the language they used in the public statements they issued to signal their position on Russia.[33] I this study I implement a form of Latent Dirichlet Allocation (LDA) for Structural Topic Models (Blei 2012; Blei, Ng, and Jordan 2003). The underlying idea behind LDA is that a document of text focused on the same topic will contain similar words, and the central assumptions is that a corpus might contain many different topics. LDA functions to uncover topic distributions across documents, and word distributions across topics.

The topic models generate through the LDA (Latent Dirichlet Allocation) method are depicted in the table (6) below. The total number of unique tokens (words) are 5,863, with a total number of documents at 300. The table of key terms shows the distribution of terms clustered around 7 topics.[34] I then distinguish between statements from withdrawing firms and statements from remaining firms. I make a hard cut off here for withdrawing firms as 'failing grades' (buying time, digging in), and withdrawing firms as 'A and B grades' (withdrawal, scaling back). The reason I exclude suspending firms from the analysis is that these firms are in limbo and tend to either move in the direction of withdrawal or digging in, and including them as withdrawing firms would skew the findings. The table shows the distribution of topics between remaining and withdrawing firms, and the graph depicts this distribution difference across topics. The remaining group has a stronger association with Topic 1 (which emphasized support for Ukraine, humanitarian efforts, and the well-being of employees). The

[33] In the appendix, I go into more detail on differences between company statements using word-embedding.

[34] In the appendix I use a Structural Topic Model (STM) to estimate the effects of the scale of withdrawal as a covariate. The results reinforce the main findings that remaining firms are more likely to make appeals to humanitarian efforts to deflect attention.

withdrawing group has a more even distribution across the topics, with notable associations with Topics 2 (business operations in Russia and Ukraine) and 4 (ongoing operations and safety). The notable finding here is that remaining firms place a greater emphasis on topics related to humanitarian aid and the well-being of their Russian employees.

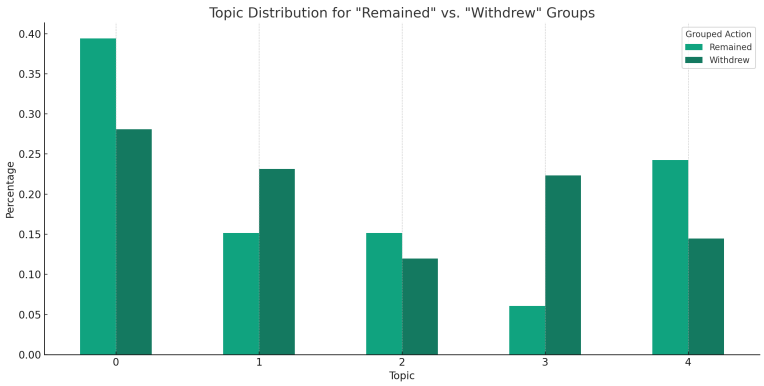
Topics 1 and 3 revolve around support for Ukraine, humanitarian efforts, and the well-being of employees and people. Topic 2 focuses on business operations in Russia and Ukraine, with mentions of the war. Topic 4 touches upon ongoing operations and safety, and topic 5 emphasizes human rights and international concerns.

**Table 6.** LDA Topic Models: Withdrawing Firm Statements

| Topic | Key Terms |
|---|---|
| Topic 1: | investments supplies businesses european impact universal state pursuant primarily corps |
| Topic 2: | investments supplies businesses local temporary addition exports corps european |
| Topic 3: | improve european hardship supplies investments invite pursuant inquiry primarily |
| Topic 4: | european invite logistical temporary rights total universal site monitoring |
| Topic 5: | companies businesses european proud support conduct pursuant accordingly |
| Topic 6: | businesses addition industry supplying course commitments sensitive billion coordination |
| Topic 7: | condemned fair human businesses reporting supplies terrible investments manages place |

**Table 7.** LDA Topic Models: Remaining Firm Statements

| Topic | Key Terms |
|---|---|
| Topic 1: | ukrainian support group colleagues bank refugees including customers euro unhcr |
| Topic 2: | decided employees diligence teams authorities distribution region situation operate suspended |
| Topic 3: | group rights human business respect situation million fund people fulfilling |
| Topic 4: | sanctions support people employees country international products use local war |
| Topic 5: | humanitarian support people foundation ukrainian efforts health medicines relief continue |
| Topic 6: | rights human business principles international respect group situation activities global |
| Topic 7: | ground reaching management security issue halted basis extended thank guidance |




### 10.3 UNGC Membership

I test cosine similarity of UNGC and UNGP membership references between remainders and withdrawers. The result below show that remainders are much

more likely to invoke membership to the UNGC. Additional I test whether firms that belong to the UNGC are more likely to withdraw from Russia. The results suggest UNGC membership has a negative but statistically insignificant effect on likelihood of withdrawal.

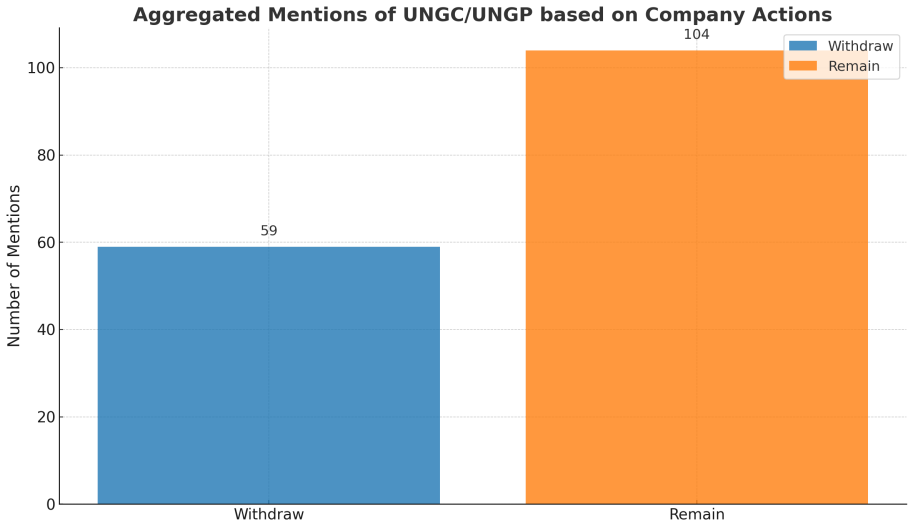



**Table 8.** Regression results: UNGC member

| | Estimate | Std. Error | t value | Pr(¿—t—) |
|---|---|---|---|---|
| (Intercept) | 0.45882 | 0.16293 | 2.816 | 0.00915 ** |
| ungc_member | -0.17646 | 0.12006 | -1.470 | 0.15362 |
| firm_size_categoryMid Cap | 0.05090 | 0.16758 | 0.304 | 0.76375 |
| firm_size_categorySmall Cap | -0.08200 | 0.23392 | -0.351 | 0.72874 |
| factor(industry)Consumer Discretionary | 0.38130 | 0.22554 | 1.691 | 0.10287 |
| factor(industry)Consumer Staples | 0.01799 | 0.20910 | 0.086 | 0.93211 |
| factor(industry)Energy | 0.21573 | 0.26591 | 0.811 | 0.42456 |
| factor(industry)Financials | 0.24235 | 0.20749 | 1.168 | 0.25340 |
| factor(industry)Industrials | 0.18530 | 0.18547 | 0.999 | 0.32697 |
| factor(industry)Information Technology | 0.54118 | 0.31840 | 1.700 | 0.10113 |
| factor(industry)Materials | -0.07743 | 0.21325 | -0.363 | 0.71947 |
| factor(industry)NGO | 0.47451 | 0.19753 | 2.402 | 0.02373 * |
| factor(industry)Utilities | 0.11764 | 0.32586 | 0.361 | 0.72101 |

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**Table 9.** Regression results: UNGC Membership Duration

| | Estimate | Std. Error | t value | Pr(¿—t—) |
|---|---|---|---|---|
| (Intercept) | $4.551 \times 10^{1}$ | $2.096 \times 10^{1}$ | 2.172 | 0.118 |
| joined_year | $-2.222 \times 10^{-2}$ | $1.037 \times 10^{-2}$ | -2.143 | 0.121 |
| firm_size_categoryMid Cap | $-2.667 \times 10^{-1}$ | $2.943 \times 10^{-1}$ | -0.906 | 0.432 |
| firm_size_categorySmall Cap | $-4.004 \times 10^{-16}$ | $2.667 \times 10^{-1}$ | 0.000 | 1.000 |
| factor(industry)Consumer Discretionary | $2.222 \times 10^{-2}$ | $2.669 \times 10^{-1}$ | 0.083 | 0.939 |
| factor(industry)Consumer Staples | $-8.000 \times 10^{-1}$ | $3.255 \times 10^{-1}$ | -2.458 | 0.091 . |
| factor(industry)Financials | $-3.556 \times 10^{-1}$ | $2.545 \times 10^{-1}$ | -1.397 | 0.257 |
| factor(industry)Industrials | $-1.556 \times 10^{-1}$ | $2.319 \times 10^{-1}$ | -0.671 | 0.550 |
| factor(industry)Materials | $-4.444 \times 10^{-1}$ | $2.675 \times 10^{-1}$ | -1.662 | 0.195 |
| factor(industry)Utilities | $-4.444 \times 10^{-1}$ | $2.900 \times 10^{-1}$ | -1.532 | 0.223 |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.1886 on 3 degrees of freedom
(977 observations deleted due to missingness)
Multiple $R^2$: 0.8136, Adjusted $R^2$: 0.2545
F-statistic: 1.455 on 9 and 3 DF, p-value: 0.4178

The findings show that UNGC member firms are no more likely to depart Russia than non-member UNGC firms. Membership duration has no effect on decisions to withdraw, offering counter-evidence to theories that suggest early adopters of regulatory initiatives may be more committed to the promotion of socially responsible norms due to relatively higher levels of ideational commitment (Finnemore and Sikkink 1998), or their prospect of comparatively large efficiency gains (Tolbert and Zucker 1983).

These findings confirm past studies showing that multi-stakeholder initiatives for governing multinational corporations do not lead to significant changes in output legitimacy, defined as "whether MSIs are effective in addressing the issues they claim to address" Scharpf 2009; Rasche, Gwozdz, et al. 2022, in part because the reputational mechanism that should induce firm compliance by pressuring the firm to alter behaviour through appeals, or by the prospect of de-

listment or non-communication, is not perceived as a credible threat by the firm.

## 11 Conclusion

So, who gives a damn about a bad reputation? The findings in this paper suggest that reputational costs induced by peer firm withdrawals within an sector influenced firm decisions to withdraw from Russia initially. This effect, however, not only declined over time, but reversed for firms that decided to leave later. Firms that remained were not immune to fears of reputational damage. Rather, these firms attempted to balance reputational costs from both directions, their Russian audience and an international consumer base, by utilizing a key set of strategies to shelter and preserve their reputation. Greater visibility and an institutional environment that allows for media transparency and mobilized collective action increased the reputational costs for firms in Russia to remain.

The findings reinforce some of the observational trends in the literature on business administration and finance that show a link between the rise of social media and reputation as an increasingly fragile and risky source of profit disruption and brand awareness. As one study of leaders at 300 major global companies pointed out, reputation was cited as the highest impact risk for companies, "...due in large measure to the rise of social media, which enables instantaneous global communications that make it harder for companies to control how they are perceived in the marketplace" (Legocki and Walker 2016, p. 1). While this paper is focused on the supply-side responses to demand, recent survey research on business decisions to withdraw or remain has found that, "Only 37% of the respondents think that leaving Russia is a pure business decision, and only 30% think that sanctions are a pure matter for the government. If a firm does not

conform to the desire to leave Russia, 66% of the respondents are willing to boycott it (exit)" (O. D. Hart, Thesmar, and Zingales 2022, p. 203). These findings suggest that consumers *do* place an emphasis on reputations for human rights that will affect how they engage with firms. The role of consumers, not just voters, in pressuring companies through reputational costs to change their behavior is a significant finding for modern warfare and the use of private sanctions in future conflicts.

The actions of early moving firms on their peers within a sub-sector had a large effect on pushing other firms to withdraw, in some cases these were firms that had more to lose, including greater revenue exposure to the Russian market, more shareholder value that was at risk, and running the high risk of making themselves a target for the Kremlin and Russian regulators to threaten to seize their assets. The costs of withdrawing from Russia were steep for some firms, for instance McDonald's incurred a a $1.4bn accounting charge. Yet, these firms still reacted by issuing statements of intent to exit Russia and go beyond the bare minimum required by sanctions. The daily time-series element of the data on issuing statements to withdraw in this paper capitalizes on the timing of decisions for this reason. These firms perhaps would have preferred more time to make a decision, make arrangements to divest, and quietly leave without staining their prior neutrality with political activism that could harm them in other politically charged regions. The fact that so many firms felt it necessary to publicly signal their intentions to depart Russia directly to their consumer base on social media platforms is suggestive of the power that reputation-stealing externalities can have on companies.

Content and text analysis of the nature of firm positions on Russia through the lens of corporate statements provides evidence that firms that remained attempted to strategically manage their reputation. While firms that withdrew effectively 'burnt the bridge' with the Russian market by issuing condemnatory and inflammatory statements, firms that remained used more cautious language that was less critical of Russia, less likely to use words like 'invaded' and more likely to focus on compliance. These remaining firms tried to deflect against the reputational costs of remaining from their international consumer base by appealing to their humanitarian actions, and leveraging their membership to UN-adjacent MSIs.

In addition, I find two pieces of evidence that corroborate past findings that are critical of the UNGC, suggesting that not only is the multi-stakeholder initiative ineffective in restraining members, but that the organization is used to leverage legitimacy and mitigate reputational costs for firms that remain. This lends support for the findings of Berliner and Prakash (2015) that the UNGC's lack of monitoring and enforcement enables member companies to shirk their obligations while using the organization as a shield for poor behavior. Preliminary evidence suggests this effect might go further; bad actors learning from their peers and joining the organization post-invasion to shore up social capital. This has not gone unnoticed by member states; in June of 2023, Vodaphone announced its withdrawal from the United Nations Global Compact in reaction to the languid behavior of the UNGC towards disciplining member states, and its obsequiousness to its powerful members that used the organization to defend their behavior, claiming in their statement, "And we do not consider it possible to remain among the initiatives of an organization that itself does not demonstrate proper responsibility and avoids using its levers of influence for the sake of

sustainable development" ("Vodafone Ukraine Leaves the UN Global Compact Network" 2023).

This paper also provides some preliminary evidence of a relationship between corporate decisions to remain in Russia and corporate philanthropic giving. While not a quantitative study of explicit amounts donated, this paper shows that, remaining firms attempting to maintain a presence in Russia are more likely to emphasize their philanthropic activity and the provision of humanitarian aid in Ukraine. This supports a thin branch of existing research on the use of philanthropy as a deflection technique to dress up politically and socially contentious decisions (C. Chen, Patten, and R. W. Roberts 2008; Du et al. 2014; Koehn and Ueng 2010). Further research is needed to determine whether this is backed up by financial data, or if there is a mismatch between discourse and action.

The findings in this paper also lend support for the role of reputational cascades as a mechanism through which reputation can affect behavior. Rather than a one-dimensional view of reputation as being history-dependent, or calculated through obscure practices across several unrelated dimensions over time (ESG), this paper argues that we should consider who the audience is, and how they cognitively evaluate reputation, when measuring the effects of reputation on firm behavior. I show that intra-industry diffusion create reputational externalities on firms; firms are more likely to withdraw quickly from Russia when their competitors also withdraw. In this way, reputation can be re-conceptualized as a group-based phenomenon decided by the average of the group a firm belongs to, rather than a history-dependent concept based on the firm's behavior in isolation. The effect of withdrawing is to move the benchmark of 'good' or 'bad'

behavior for consumers to evaluate, and to effectively cast a spotlight on the competitors of the firm that left. While firm size, exposure to the Russian market, influence of sanctions, and shareholder losses surely play a role in influencing firm behavior, the findings in this paper suggest that reputational concerns were a primary motivator. The paper also contributes to the literature on the effects of naming-and-shaming in International Relations, with particular application to firms.

One key limitation in this paper is that firms are differentially affected by the imposition of sanctions depending on their use of trade finance instruments; firms that strongly depend on such instruments will be greatly affected by such sanctions. Evidence also suggested that firms with prior experience operating within the sanctioned country are more adapt and even immune to the effects of sanctions. Such firms become specialized in serving 'crisis countries' and can withstand sanctions. Secondly, an emerging trend has seen firms avoid sanctions by exporting indirectly via neighboring countries (Crozet et al. 2021). Future research should investigate the impact of sanctions on financial instruments, and how firms used neighboring countries to continue operations. A second limitation A third limitation is the extent to which these findings could be generalized to other contexts given the specific features of this situation, however the online backlash against company support for particular sides in the Israel-Palestine war and calls for boycotts, suggests similar scenarios are possible in the future.

Another avenue of future research is examining who replaced these withdrawing companies. One of the sheltering strategies adopted by Russia to weather the storm of economic sanctions was to pivot trade - both export destination and import source - from Western countries towards certain Asian economies.

Russian-Chinese trade grew 29 percent in 2022, and 39 percent in 2023, projected to reach US$237 by the end of the year. In 2022, Russia's trade with Turkey increased 87 percent, 68 percent with the United Arab Emirates, and a whopping 205 percent with India. Asia has counterbalanced the loss of the European oil market, with India, China, Turkey and the Gulf states perfectly replacing their European counterparts. The corporate exodus out of Russia by Western companies has been filled by Chinese firms, now constituting 40 percent of Russia's car sales, and 70 per cent of smartphone sales. Chinese and Indian firms dominate the 'scaling up' category, are less likely to issue media statements, almost no Chinese or Indian firm issues a social media statement, and any media reference made explicitly omits the terms 'invasion'. Interestingly, Didi, the Chinese company with a ride-hailing unit in Russia, reversed its decision to withdraw from Russia. On February 21st 2022, it cited "market changes and other challenges" as the reasons for its withdrawal. Within five days the company reversed its decision following consumer backlash on Weibo, with one commentator exclaiming, "Didi is a running dog [for the US]. Let's delete the app and let it get out of China". What happened in China was the opposite of what we observed in Western countries; companies risked their reputations if they withdrew early. Future research should explore this duality further.

The foundations of neoclassical economics are grounded in the idea that firms are always seeking to maximise profits. Firm behavior, and consumer reactions during Russia's invasion of Ukraine, suggest that perhaps treating these entities as a-political actors that have limited impact on geopolitics does not reflect the new reality of these agents. Given the unprecedented withdrawal of firms from Russia, this case sheds light on the way in which business actors are adapting, and creating changes within, the geopolitical environment. Where economic

statecraft becomes an increasingly useful instrument to fight international conflicts. This paper contributes to the debate on the role of economic statecraft, the role of firms as agents in international relations, and how reputational concerns may be an effective tool of altering MNC's behavior. These findings have broader implications for how states manage their own reputations in relation to signaling disapproval of another states behavior to appease the international community of states they belong to, while simultaneously using careful language to allow for a reputational management amongst different audiences.

**Keywords:** Sanctions · Reputation · Ukraine · Russia · Economic Statecraft.

## References

Addo, Michael K (2014). "The reality of the United Nations guiding principles on business and human rights". In: *Human Rights Law Review* 14.1, pp. 133–147.

Ahmed, Daniyal et al. (2023). "Are ESG Ratings Informative About Companies' Socially Responsible Behaviors Abroad? Evidence from the Russian Invasion of Ukraine". In: *Accountability in a Sustainable World Quarterly, Forthcoming*.

Amer, Estefania (2018). "The penalization of non-communicating UN Global Compact's companies by investors and its implications for this initiative's effectiveness". In: *Business & Society* 57.2, pp. 255–291.

Baker, Andrew C, David F Larcker, and Charles CY Wang (2022). "How much should we trust staggered difference-in-differences estimates?" In: *Journal of Financial Economics* 144.2, pp. 370–395.

Baldwin, David A (2020). *Economic statecraft: New edition*. Princeton University Press.

Beck, Nathaniel, Jonathan N Katz, and Richard Tucker (1998). "Taking time seriously: Time-series-cross-section analysis with a binary dependent variable". In: *American Journal of Political Science* 42.4, pp. 1260–1288.

Bénabou, Roland and Jean Tirole (2010). "Individual and corporate social responsibility". In: *Economica* 77.305, pp. 1–19.

Berliner, Daniel and Aseem Prakash (2015). ""Bluewashing" the Firm? Voluntary Regulations, Program Design, and Member Compliance with the U nited N ations G lobal C ompact". In: *Policy Studies Journal* 43.1, pp. 115–138.

Bernhagen, Patrick, Neil J Mitchell, and Marianne Thissen-Smits (2013). "Corporate citizens and the UN Global Compact: Explaining cross-national variations in turnout". In: *Business and Politics* 15.1, pp. 63–85.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch (1992). "A theory of fads, fashion, custom, and cultural change as informational cascades". In: *Journal of political Economy* 100.5, pp. 992–1026.

Blei, David M (2012). "Probabilistic topic models". In: *Communications of the ACM* 55.4, pp. 77–84.

Blei, David M, Andrew Y Ng, and Michael I Jordan (2003). "Latent dirichlet allocation". In: *Journal of machine Learning research* 3.Jan, pp. 993–1022.

Bromley, Patricia and Walter W Powell (2012). "From smoke and mirrors to walking the talk: Decoupling in the contemporary world". In: *Academy of Management annals* 6.1, pp. 483–530.

Bruno, Kenny and Joshua Karliner (2002). "The UN's global compact, corporate accountability and the Johannesburg Earth Summit". In: *Development* 45.3, pp. 33–38.

C. Chen, Jennifer, Dennis M Patten, and Robin W Roberts (2008). "Corporate charitable contributions: A corporate social performance or legitimacy strategy?" In: *Journal of Business Ethics* 82, pp. 131–144.

Carlos, W Chad and Ben W Lewis (2018). "Strategic silence: Withholding certification status as a hypocrisy avoidance tactic". In: *Administrative Science Quarterly* 63.1, pp. 130–169.

Castioni, Piergiorgio et al. (2022). "The voice of few, the opinions of many: evidence of social biases in Twitter COVID-19 fake news sharing". In: *Royal Society Open Science* 9.10, p. 220716.

Cengiz, Doruk et al. (2019). "The effect of minimum wages on low-wage jobs". In: *The Quarterly Journal of Economics* 134.3, pp. 1405–1454.

Chen, Yuyu and David Y Yang (2019). "The impact of media censorship: 1984 or brave new world?" In: *American Economic Review* 109.6, pp. 2294–2332.

Coen, David, Kyle Herman, and Tom Pegram (2022). "Are corporate climate efforts genuine? An empirical analysis of the climate 'talk–walk'hypothesis". In: *Business Strategy and the Environment* 31.7, pp. 3040–3059.

Coffee, Patrick and Lindsay Rittenhouse (Mar. 15, 2022). "Most ad agencies haven't withdrawn from Russia, and employees are asking why". In: *Business Insider*. URL: `https://www.businessinsider.com/advertising-agencies-havent-left-russia-employees-are-asking-why-2022-3` (visited on 03/20/2022).

Cook, Scott J, Jude C Hays, and Robert J Franzese (2023). "STADL Up! The Spatiotemporal Autoregressive Distributed Lag Model for TSCS Data Analysis". In: *American Political Science Review* 117.1, pp. 59–79.

Crozet, Matthieu et al. (2021). "Worth the pain? Firms' exporting behaviour to countries under sanctions". In: *European Economic Review* 134, p. 103683.

Cyert, Richard M, James G March, et al. (1963). "A behavioral theory of the firm". In: *Englewood Cliffs, NJ* 2.4, pp. 169–187.

Davis, Christina L, Jialu Li, and Sayumi Miyano (2023). "Peer Conformity and Competition: How Business Managers Evaluate Firm Withdrawals from Russia". In.

DellaVigna, Stefano, John A List, and Ulrike Malmendier (2012). "Testing for altruism and social pressure in charitable giving". In: *The quarterly journal of economics* 127.1, pp. 1–56.

Deva, Surya and David Bilchitz (2013). *Human rights obligations of business: beyond the corporate responsibility to respect?* Cambridge University Press.

DiMaggio, Paul J and Walter W Powell (1983). "The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields". In: *American sociological review*, pp. 147–160.

Drempetic, Samuel, Christian Klein, and Bernhard Zwergel (2020). "The influence of firm size on the ESG score: Corporate sustainability ratings under review". In: *Journal of business ethics* 167, pp. 333–360.

Du, Xingqiang et al. (2014). "Corporate environmental responsibility in polluting industries: Does religion matter?" In: *Journal of Business Ethics* 124, pp. 485–507.

Durante, Ruben and Ekaterina Zhuravskaya (2018). "Attack when the world is not watching? US news and the Israeli-Palestinian conflict". In: *Journal of Political Economy* 126.3, pp. 1085–1133.

Dyck, Alexander, Natalya Volchkova, and Luigi Zingales (2008). "The corporate governance role of the media: Evidence from Russia". In: *The Journal of Finance* 63.3, pp. 1093–1135.

Endres, Kyle and Costas Panagopoulos (2017). "Boycotts, buycotts, and political consumerism in America". In: *Research & Politics* 4.4, p. 2053168017738632.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova (2020). "Social media and protest participation: Evidence from Russia". In: *Econometrica* 88.4, pp. 1479–1514.

Finnemore, Martha and Kathryn Sikkink (1998). "International norm dynamics and political change". In: *International organization* 52.4, pp. 887–917.

Firth, John (1957). "A synopsis of linguistic theory, 1930-1955". In: *Studies in linguistic analysis*, pp. 10–32.

Fiss, Peer C and Edward J Zajac (2006). "The symbolic management of strategic change: Sensegiving via framing and decoupling". In: *Academy of management journal* 49.6, pp. 1173–1193.

Gayo-Avello, Daniel (2012). "" I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper"–A Balanced Survey on Election Prediction using Twitter Data". In: *arXiv preprint arXiv:1204.6441*.

Ginzel, Linda E, Roderick M Kramer, and Robert I Sutton (2004). "Organizational impression management as a reciprocal influence process: The neglected role of the organizational audience". In: *Organizational identity*, pp. 223–261.

Guriev, Sergei, Nikita Melnikov, and Ekaterina Zhuravskaya (2021). "3g internet and confidence in government". In: *The Quarterly Journal of Economics* 136.4, pp. 2533–2613.

Hamid, Uzma and Oliver Johner (2010). "The United Nations Global Compact communication on progress policy: Origins, trends and challenges". In: *The United Nations global compact: Achievements, trends and challenges*, pp. 265–280.

Hamilton, William L, Jure Leskovec, and Dan Jurafsky (2016). "Diachronic word embeddings reveal statistical laws of semantic change". In: *arXiv preprint arXiv:1605.09096*.

Hart, Oliver, David Thesmar, and Luigi Zingales (2023). "Private sanctions". In: *Economic Policy*, eiad041.

Hart, Oliver D, David Thesmar, and Luigi Zingales (2022). *Private sanctions*. Tech. rep. National Bureau of Economic Research.

Jansen, Bernard J et al. (2009). "Twitter power: Tweets as electronic word of mouth". In: *Journal of the American society for information science and technology* 60.11, pp. 2169–2188.

Jervis, Robert (2017). *How statesmen think: The psychology of international politics*. Princeton University Press.

Kahneman, Daniel and Amos Tversky (1973). "On the psychology of prediction." In: *Psychological review* 80.4, p. 237.

Katz, Michael L and Carl Shapiro (1994). "Systems competition and network effects". In: *Journal of economic perspectives* 8.2, pp. 93–115.

Kauffman, Robert J and Bin Wang (2007). *Duration of internet firms: A semi-parametric Cox and baysian survival analysis*. Tech. rep. Citeseer.

Koehn, Daryl and Joe Ueng (2010). "Is philanthropy being used by corporate wrongdoers to buy good will?" In: *Journal of Management & Governance* 14, pp. 1–16.

Krieg, Kimberly S and John Li (2021). "A review of corporate social responsibility and reputational costs in the tax avoidance literature". In: *Accounting Perspectives* 20.4, pp. 477–542.

Legocki, Kimberly and Kristen Walker (2016). "This is OUTRAGEOUS!: Understanding Consumer Anger and Brand Reputation on Social Media". In.

Lerner, Michael and Iain Osgood (2023). "Across the Boards: Explaining Firm Support for Climate Policy". In: *British Journal of Political Science* 53.3, pp. 934–957.

Li, Lambert Zixin and Sarah A Soule (2022). "The Power of Public Confession: Mobilization and Reputation Effects of Disclosing Socially Irresponsible Performance". In: *Available at SSRN 4278369*.

Li, Zhao and Richard W Disalvo (2023). "Can stakeholders mobilize businesses for the protection of democracy? Evidence from the US Capitol insurrection". In: *American Political Science Review* 117.3, pp. 1130–1136.

Lieberman, Marvin B and David B Montgomery (1998). "First-mover (dis) advantages: retrospective and link with the resource-based view". In: *Strategic management journal* 19.12, pp. 1111–1125.

Lietuva, Orlen (2022). "THE ONLY, THE LARGEST, THE BIGGEST - these are the words describing us". In: *Linkedin*. URL: `https://www.linkedin.com/posts/public-company-orlen-lietuva-_ab-orlen-lietuva-labai-glaud%C5%BEiai-susijusi-activity-7026942775178801152-Yes_?utm_source=share&utm_medium=member_desktop` (visited on 12/20/2022).

Lim, Alwyn and Kiyoteru Tsutsui (2012). "Globalization and commitment in corporate social responsibility: Cross-national analyses of institutional and political-economy effects". In: *American Sociological Review* 77.1, pp. 69–98.

Lins, Karl V, Henri Servaes, and Ane Tamayo (2017). "Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis". In: *the Journal of Finance* 72.4, pp. 1785–1824.

Lopez-de-Silanes, Florencio, Joseph A McCahery, and Paul C Pudschedl (2020). "ESG performance and disclosure: A cross-country analysis". In: *Singapore Journal of Legal Studies*, pp. 217–241.

Lu, Fangzhou, Lei Huang, and Sixuan Li (2022). "Sanctions and Social Capital: Evidence from the Russian Invasion of Ukraine". In: *Available at SSRN 4108129*.

Lyon, Thomas P and A Wren Montgomery (2013). "Tweetjacked: The impact of social media on corporate greenwash". In: *Journal of business ethics* 118, pp. 747–757.

Manacorda, Marco and Andrea Tesei (2020). "Liberation technology: Mobile phones and political mobilization in Africa". In: *Econometrica* 88.2, pp. 533–567.

McDonnell, Mary-Hunter and Brayden King (2013). "Keeping up appearances: Reputational threat and impression management after social movement boycotts". In: *Administrative science quarterly* 58.3, pp. 387–419.

Meier, Jean-Marie et al. (2023). "Do Consumers Care About ESG? Evidence from Barcode-Level Sales Data". In.

Mena, Sébastien and Guido Palazzo (2012). "Input and output legitimacy of multi-stakeholder initiatives". In: *Business Ethics Quarterly* 22.3, pp. 527–556.

Mikolov, Tomas et al. (2013). "Efficient estimation of word representations in vector space". In: *arXiv preprint arXiv:1301.3781*.

Miura, Satoshi and Kaoru Kurusu (2015). "Why do companies join the United Nations Global Compact? The case of Japanese signatories". In: *Corporate social responsibility in a globalizing world*, pp. 286–320.

Monin, Benoıt, Pamela J Sawyer, and Matthew J Marquez (2008). "The rejection of moral rebels: resenting those who do the right thing." In: *Journal of personality and social psychology* 95.1, p. 76.

Mulder, Nicholas (2022). *The economic weapon: The rise of sanctions as a tool of modern war*. Yale University Press.

Mylovanov, Tymofiy et al. (2023). "How to Identify Foreign Business in Russia and What are the Key Issues of Creating and Keeping a Full List of the Largest Foreign Companies in Russia". In: *Available at SSRN*.

Nilsson, Eva and Martin Fougère (2024). "Global South States' Views on Building Partnerships with Corporations: An Agonistic Struggle in the UN and Beyond". In: *Journal of Business Ethics*, pp. 1–19.

Ojiambo, Sanda (May 21, 2022). "Ukraine Crisis: Message to the UN Global Compact Stakeholders and Participants by Sanda Ojiambo". In: *The UN Global Compact*. URL: https://www.globalcompactnetwork.org/en/news/ungcn-italy-news/2105-ukraine-crisis-message-to-the-un-global-compact-stakeholders-and-participants-by-sanda-ojiambo-2.html.

Ottoni-Wilhelm, Mark, Lise Vesterlund, and Huan Xie (2017). "Why do people give? Testing pure and impure altruism". In: *American Economic Review* 107.11, pp. 3617–3633.

Pajuste, Anete and Anna Toniolo (2022). "Corporate Response to the War in Ukraine: Stakeholder Governance or Stakeholder Pressure?" In: *Available at SSRN 4183604*.

Papacharissi, Zizi (2015). *Affective publics: Sentiment, technology, and politics*. Oxford University Press.

Pennington, Jeffrey, Richard Socher, and Christopher D Manning (2014). "Glove: Global vectors for word representation". In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543.

Petroleum, British (2022). *bp's position in Russia. bp's position in Russia*. Accessed = 2022-11-30. URL: https://www.bp.com/en/global/corporate/news-and-insights/press-releases/bps-position-in-russia.html.

Raghunandan, Aneesh and Shiva Rajgopal (2022). "Do ESG funds make stakeholder-friendly investments?" In: *Review of Accounting Studies* 27.3, pp. 822–863.

Rasche, Andreas (2009). "Toward a model to compare and analyze accountability standards–The case of the UN Global Compact". In: *Corporate Social Responsibility and Environmental Management* 16.4, pp. 192–205.

Rasche, Andreas, Wencke Gwozdz, et al. (2022). "Which firms leave multi-stakeholder initiatives? An analysis of delistings from the United Nations Global Compact". In: *Regulation & Governance* 16.1, pp. 309–326.

Rasche, Andreas, Sandra Waddock, and Malcolm McIntosh (2013). "The United Nations global compact: Retrospect and prospect". In: *Business & Society* 52.1, pp. 6–30.

*Responsible business conduct implications of Russia's invasion of Ukraine* (Mar. 28, 2023). https://www.oecd.org/ukraine-hub/policy-responses/responsible-business-conduct-implications-of-russia-s-invasion-of-ukraine-f222a4d1/.

Rheault, Ludovic and Christopher Cochrane (2020). "Word embeddings for the analysis of ideological placement in parliamentary corpora". In: *Political Analysis* 28.1, pp. 112–133.

Roberts, Margaret E, Brandon M Stewart, and Dustin Tingley (2016). "Navigating the local modes of big data". In: *Computational social science* 51.

Roberts, Margaret E, Brandon M Stewart, Dustin Tingley, et al. (2014). "Structural topic models for open-ended survey responses". In: *American journal of political science* 58.4, pp. 1064–1082.

Rodriguez, Pedro L and Arthur Spirling (2022). "Word embeddings: What works, what doesn't, and how to tell the difference for applied research". In: *The Journal of Politics* 84.1, pp. 101–115.

Rodriguez, Pedro L, Arthur Spirling, and Brandon M Stewart (2023). "Embedding regression: Models for context-specific description and inference". In: *American Political Science Review*, pp. 1–20.

Ruggie, John (2008). "Protect, respect and remedy: A framework for business and human rights". In: *Innovations: Technology, Governance, Globalization* 3.2, pp. 189–212.

— (2011). "Report of the special representative of the secretary-general on the issue of human rights and transnational corporations and other business enterprises: Guiding principles on business and human rights: implementing the united nations 'protect, respect and remedy'framework". In: *Netherlands Quarterly of Human Rights* 29.2, pp. 224–253.

Sanders, Astrid (2014). "The impact of the'Ruggie framework'and the United Nations guiding principles on business and human rights on transnational human rights litigation". In.

Scharpf, Fritz W (2009). "Legitimacy in the multilevel European polity". In: *European Political Science Review* 1.2, pp. 173–204.

Schelling, Thomas C (1980). *The Strategy of Conflict: with a new Preface by the Author*. Harvard university press.

Schembera, Stefan (2018). "Implementing corporate social responsibility: Empirical insights on the impact of the UN Global Compact on its business participants". In: *Business & Society* 57.5, pp. 783–825.

Sethi, S Prakash and Donald H Schepers (2014). "United Nations global compact: The promise–performance gap". In: *Journal of Business Ethics* 122, pp. 193–208.

Simon, Scott (n.d.). *Russian law bans journalists from calling conflict a 'war' or an 'invasion'*. NPR. URL: https://www.npr.org/2022/03/05/1084729579/russian-law-bans-journalists-from-calling-ukraine-conflict-a-war-or-an-invasion#:~:text=Journalists%20could%20be%20jailed%20for,the%20safety%20of%20their%20employees.

Simons, John (2022). *Is Wall Street Doing Enough to Ostracize Russia? "That's Not Our Job," Says Goldman Sachs CEO. Is Wall Street Doing Enough to Ostracize Russia? "That's Not Our Job," Says Goldman Sachs CEO*. Accessed = 2022-03-30. URL: https://time.com/6157107/david-solomon-goldman-sachs-ceo-interview/.

Slinko, Irina, Evgeny Yakovlev, and Ekaterina Zhuravskaya (2005). "Laws for sale: evidence from Russia". In: *American Law and Economics Review* 7.1, pp. 284–318.

Slovic, Paul (1975). "Choice between equally valued alternatives." In: *Journal of Experimental Psychology: Human perception and performance* 1.3, p. 280.

Slovic, Paul and Douglas MacPhillamy (1974). "Dimensional commensurability and cue utilization in comparative judgment". In: *Organizational Behavior and Human Performance* 11.2, pp. 172–194.

Solis, Jonathan A and Philip D Waggoner (2021). "Measuring media freedom: An Item Response Theory analysis of existing indicators". In: *British Journal of Political Science* 51.4, pp. 1685–1704.

Sonnenfeld, Jeffrey, Tymofiy Mylovanov, et al. (2023). "The Russian Business Retreat-How the Ratings Measured Up One Year Later". In: *Available at SSRN 4343547*.

Sonnenfeld, Jeffrey, Steven Tian, et al. (2022). "It pays for companies to leave Russia". In: *Available at SSRN* 4112885.

Sorkin, Andrew (June 10, 2013). "Tech Companies Tread Lightly in Statements on U.S. Spying". In: *The New York Times*. URL: https://archive.nytimes.com/dealbook.nytimes.com/2013/06/10/tech-companies-tread-lightly-in-statements-on-u-s-spying/ (visited on 05/08/2023).

Soule, Sarah A (2018). "Social movements and their impact on business and management". In: *Oxford Research Encyclopedia of Business and Management*.

Therneau, Terry M (1997). "Extending the Cox model". In: *Proceedings of the first Seattle symposium in biostatistics: survival analysis*. Springer, pp. 51–84.

Thrall, Calvin (2021). "Public-private governance initiatives and corporate responses to stakeholder complaints". In: *International Organization* 75.3, pp. 803–836.

Tolbert, Pamela S and Lynne G Zucker (1983). "Institutional sources of change in the formal structure of organizations: The diffusion of civil service reform, 1880-1935". In: *Administrative science quarterly*, pp. 22–39.

"Vodafone Ukraine Leaves the UN Global Compact Network" (Dec. 6, 2023). In: *American Chamber of Commerce Ukraine*. URL: https://chamber.ua/news/vodafone-ukraine-leaves-the-un-global-compact-network/ (visited on 12/08/2023).

Wellhausen, Rachel L and Boliang Zhu (2024). *Exiting Russia*. Tech. rep. Working paper.

Werner, Timothy (2015). "Gaining access by doing good: The effect of sociopolitical reputation on firm participation in public policy making". In: *Management Science* 61.8, pp. 1989–2011.

Westphal, James D and Melissa E Graebner (2010). "A matter of appearances: How corporate leaders manage the impressions of financial analysts about the conduct of their boards". In: *Academy of Management Journal* 53.1, pp. 15–44.

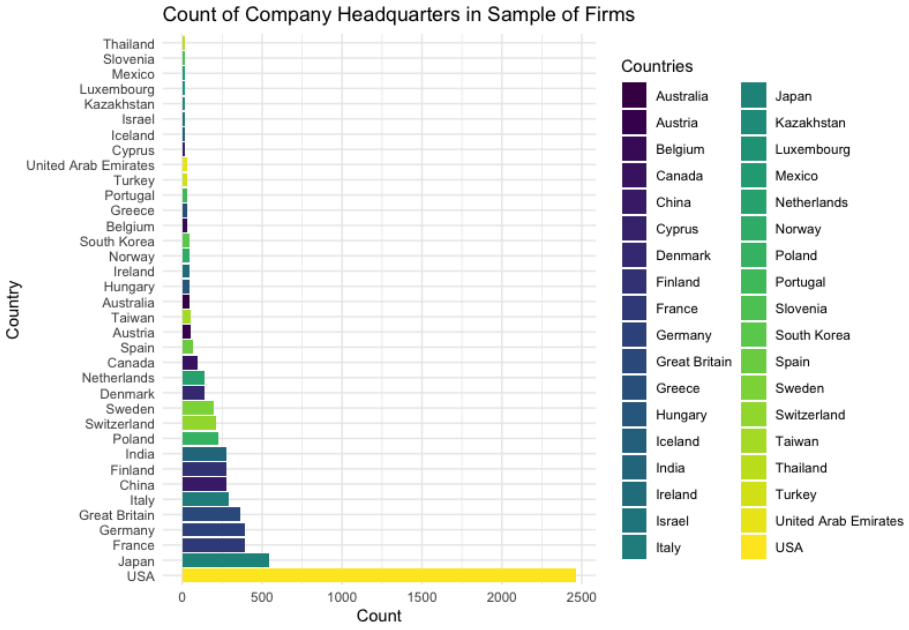
Wilson, Woodrow (1923). *Woodrow Wilson's case for the League of Nations*. Princeton: Princeton University Press.

# 12 Appendix

# 13 Descriptive Statistics

My sample of firms relies on two datasets with some supplementary inclusions of firms that I added. The bulk of this collection, however, was carried out by the Yale Businesses School team Sonnenfeld, Tian, et al. 2022; Sonnenfeld, Mylovanov, et al. 2023, and the KSE team (Mylovanov et al. 2023). I chose to use both datasets because while the Yale dataset was thorough, the KSE dataset included indicators that the Yale dataset omitted, and indicators that cannot be found on existing financial databases (including financial exposure to Russia). Here I include some more information on the methods used by both teams.

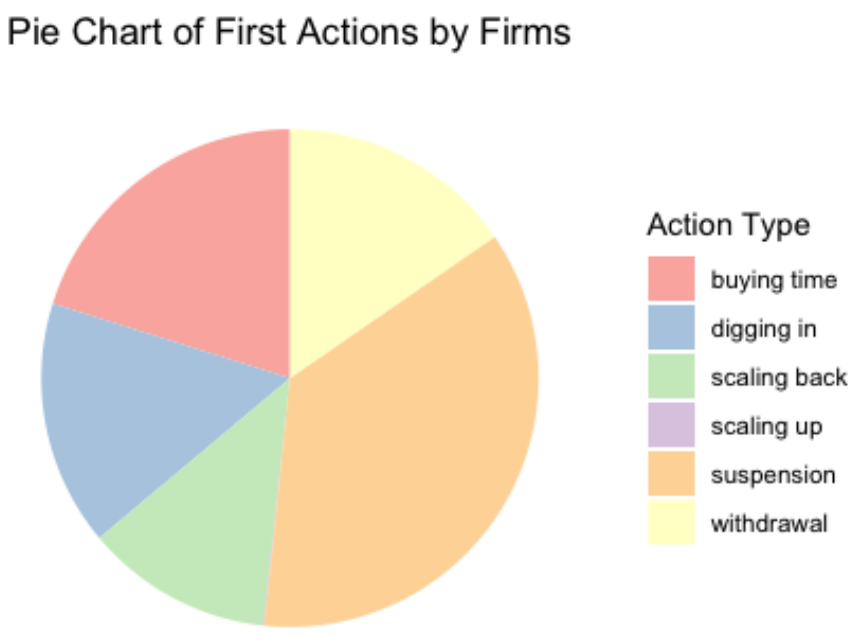
The break-down of company headquarters in the sample is detailed below - the United States is the dominant country of headquarters for companies. This is unsurprising given the context in which the data was collected, combined with the number of multi-national corporations of US origin.

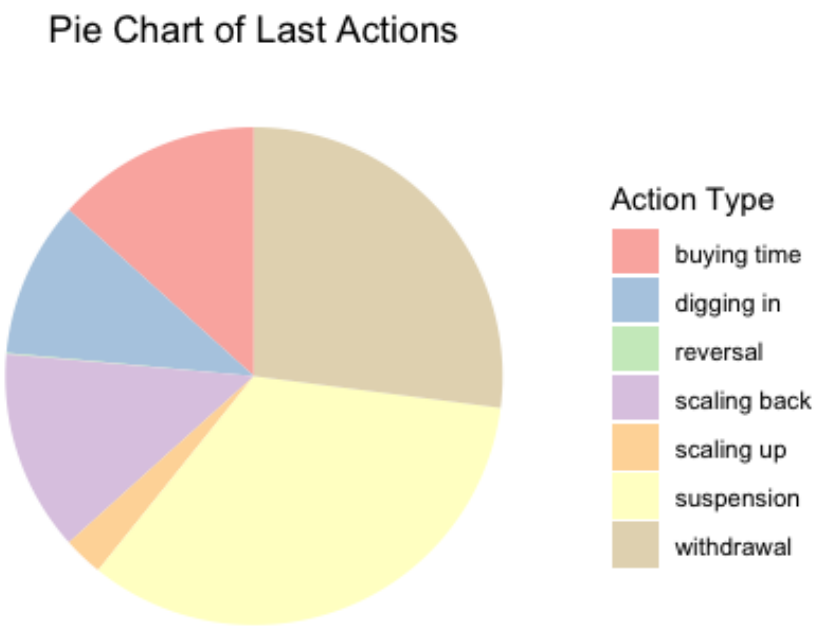



## 13.1 Coding Actions

My dataset departs substantially from Sonnenfeld, Tian, et al. 2022 or Mylovanov et al. 2023 in terms of identifying a company's status. For instance, a firm that might be listed as 'withdrawn' under the Yale database or the KSE database may be listed as 'buying time' and then 'suspended' in my database. The reason for this is that both of these databases are updating when firms have withdrawn - in some cases because these firms have reached out to the authors to signal they have withdrawn and that they would like to receive an 'A' grade. The reaction of companies to the grades they were assigned is suggestive of just how concerned these companies were with the reputational effects of remaining. Take for instance Krispy Kreme, who, in response to a Tweet that the company was still operating in Russia, Krispy Kreme replied, "We are completely winding down our franchisee's business in Russia and expect all shops to be closed in the very near term. We've shared this with Yale, which has moved Krispy Kreme to

its “A” grade of companies withdrawing from Russia.” Krispy Kreme was not the only company that adjusted its behaviour in response to the public backlash and moved from digging in to withdrawal, nor was it the only company that chose to use it’s public platform to reach as large an audience as possible that they had joined the A-grade listers. Unlike KSE and Sonnenfeld, what is meaningful for my reputation mechanism is tracking when companies issued statements on their actions, and marking that action at that date. By allowing all companies to have ’withdrawn’ statuses, it conflates two companies as essentially having behaved the same - ’withdrawn’ - when one company withdrew February 25th and the other withdrew July 6th. The two plots below illustrate the split of company actions in ’first actions’ issued versus the last recorded date of actions in August 2022.


Pie Chart of First Actions by Firms
Action Type
buying time
digging in
scaling back
scaling up
suspension
withdrawal




## 14 Data Collection

### 14.1 GeoRev: Geographical revenue exposure to Russia

Ideally data for geographical revenue exposure for companies would be collected using FactSet data. However, subscriptions to FactSet at Oxford are only accessible through programs with the Said Business School, not the Department of Politics and International Relations. I managed to obtain some of this data when visiting at Cornell, and found it unsatisfactory for the sample of companies used in this paper. I therefore manually collected this data by trolling through company financial statements in 2022 and quarterly reports. Many publicly-traded companies were eager to remind shareholders of their low exposure to the Rus-

sian market if they chose to withdraw/or if they needed to pause operations, while other companies who withdrew were eager to either publicly express their dependence on Russia and therefore demonstrate their moral goodwill in leaving Russia (take for instance the Brewing company 'InBev' that stated it had the largest share of revenue in the Russian market compared to its Western brewery peers at 11%.), to help explain revenue drops to shareholders and reassure them it was temporary and not a trend, or to justify to shareholders why the company would be remaining in Russia. I also used reports from the US Securities and Exchange Commission. I supplemented this with KSE data that included information about revenue exposure in Russia in 2021, and in 2022.

### 14.2 ESG as Reputation

### 14.3 Reputation and social pressure

Existing research has found ESG scores as poor predictors of firm decisions to withdraw Ahmed et al. 2023; Lu, Huang, and S. Li 2022. Past research on ESG as a signal of socially conscious corporate behaviour and actual ESG-related output has had mixed findings; some studies find a negative relationship between ESG scores and firm compliance over carbon emissions or meaningful sustainability initiatives Raghunandan and Rajgopal 2022; Drempetic, Klein, and Zwergel 2020; Lopez-de-Silanes, McCahery, and Pudschedl 2020.

Three main weaknesses in existing research preclude a more robust understanding of the role of reputation in inducing firm compliance, timing of action, and how firms that remain strategically avoid or mitigate reputational costs. First, reputational concerns on firm corporate responsibility have largely been focused on environmental issues, given the lower visibility of peripheral conflicts amongst Western audiences, and previous large-scale conflicts occurred at a time

when limited agency was assigned to private actors. The politicization of corporations is a relatively recent phenomenon, with the "firms as profit maximizers" the dominant discourse until recently. Relatedly, consumers had few platforms to foster collective action against multiple corporations. Informational asymmetries were higher, and the costs of mobilizing in the absence of an online platform were also higher. Second, reputational cascades have been overlooked in favour of informational cascades – the idea that a corporation could be making a costly movement in reaction to competing firm behaviour has been framed as firm A leaves and sends a signal to firm B that firm A has information about the benefits of leaving/costs of staying, leading firm B to imitate firm A. This idea that firm B is only moving because firm A supposedly has private information on the financial costs of remaining overlooks a related, but distinct, phenomenon. That firm B has 'moved the needle' of appropriate behaviour and firm A must now mimic firm B or risk the consequences.

Lastly, the analysis on reputation often assume that both reputational costs are homogenous, and that some firms are insensitive to such costs. MNC's often face multiple audiences, each of which they must maintain or protect reputations with; at some points in time, as in the Russian invasion of Ukraine, the firms reputation with these audiences can come into conflict. In this situation, some firms choose to throw their proverbial eggs into one audience, and protect one reputation, while other firms attempt to balance both at once. In examining the nature of the behaviour in how firms withdrew or remained in Russia, I am able to explore the strategies employed by firms who behave badly, and anticipate backlash.

The business and finance literature often uses ESG scores as a proxy for firm reputation. This measurement choice is based on the intuition that firms build up social capital over time, and that this measurement represents the history of firm behaviour; firms with higher scores have more to lose than firms with lower scores. In some ways ESG is a history-dependent definition of reputation, but as previously discussed, interpretation of ESG as being this are fundamentally flawed. ESG scores are computed annually, not over decades. They're often a combined score on multiple dimensions, and bad behaviour on one category can be cancelled out by good behaviour on another. They're tallied over the course of a year, not in response to crisis. In contrast, the IR literature often treats state resolve during a crisis as a marker of a state's reputation; rather than the state building up a measure of resolve in non-crisis situations over the course of a year. Audiences are primed to care about the current cohort of firm behaviour in response to the war, rather than reflecting on firm's past behaviour in different situations. Thus, in the case of a genuine geopolitical crisis, it makes little sense to use ESG as a marker of firm reputation, rather than evaluating the situational features and pressure exerted on firms by their competitors which may generate reputational costs. I test whether there is a relationship between raw ESG scores amongst firms and whether or not the firm chooses to withdraw, using a logit model with country and industry fixed effects. The results in table 3 show a positive but not statistically significant relationship beteen ESG and withdrawal. As an added robustness check I construct a 'passing grade' variable that expands to categories 'suspension' and 'scaling back'.

### 14.4 Executive Compensation

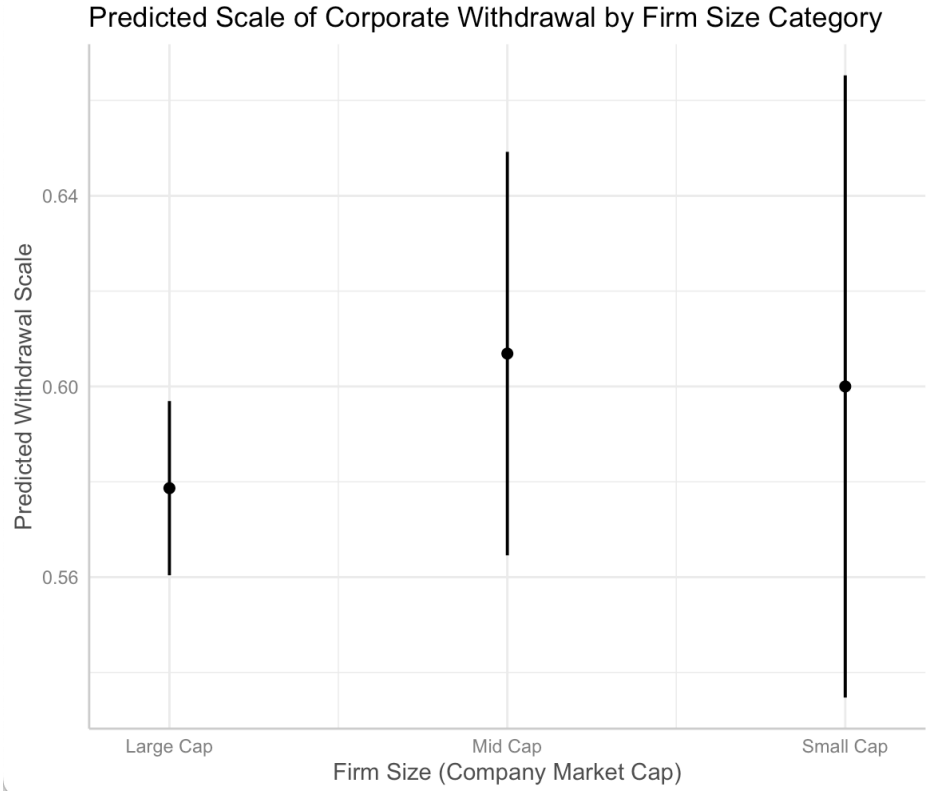
Another source of evidence to support the idea that firms were motivated by fear of reputational penalties and reputational premiums from withdrawing early would be whether executives in the firm are rewarded for socially conscious behaviour by investors or stakeholders. If there is a premium placed by investors or stakeholders on firm withdrawal, then individual managers should be more inclined to push for withdrawal for personal economic gain, as well as believing there will be little lost in investment/company shares over time. To test this I collect information on executive compensation policies in firms, since this information is not available for all the observations in my data, I am only able to collect information on 664 firms, with 42 percent of these firms having explicitly no executive compensation for ESG performance, and 58 percent offering compensation. I regress executive compensation on both a dummy variable of whether or not the firm chose to withdraw as their 'first action', and as a robustness check, included a weighted variable of the extent of withdrawal (the grade assigned to the firm), with two-way fixed-effects on country of company headquarters and the industry sector, and also cluster standard errors by country, and control for firm size. The table below displays the findings; there is a positive but statistically insignificant relationship between firms that have executive compensation policies in place and whether or not the firm chose to withdraw as their first action. Observations were dropped due to data availability.

**Table 10.** Regression Results

| | *Dependent variable:* | |
|---|---|---|
| | Model 1 (1) | Model 2 (2) |
| ESG Raw | 0.0005 (0.001) | |
| ESG Executive Compensation | | 0.015 (0.019) |
| Firm Size Category | −0.014 (0.023) | −0.016 (0.022) |
| Industry | −0.051 (0.036) | −0.060* (0.034) |
| Country | −0.302*** (0.098) | −0.306*** (0.098) |
| factor(industry)Consumer Discretionary | −0.298*** (0.094) | −0.299*** (0.094) |
| factor(industry)Consumer Staples | −0.423*** (0.096) | −0.423*** (0.096) |
| factor(industry)Energy | −0.329*** (0.098) | −0.333*** (0.098) |
| factor(industry)Financials | −0.322*** (0.096) | −0.322*** (0.096) |
| factor(industry)Health Care | −0.518*** (0.100) | −0.518*** (0.100) |
| factor(industry)Industrials | −0.325*** (0.094) | −0.325*** (0.094) |
| factor(industry)Information Technology | −0.288*** (0.095) | −0.289*** (0.096) |
| factor(industry)Manufacturing | −0.378* (0.223) | −0.375* (0.223) |
| factor(industry)Materials | −0.390*** (0.097) | −0.390*** (0.097) |
| factor(industry)Real Estate | −0.174 (0.118) | −0.170 (0.118) |
| factor(industry)Utilities | −0.281** (0.110) | −0.285** (0.111) |
| factor(country)Austria | −0.295* (0.176) | −0.288 (0.176) |
| factor(country)Belgium | −0.302 (0.205) | −0.304 (0.205) |
| factor(country)Brazil | 0.167 (0.204) | 0.170 (0.204) |
| factor(country)Canada | 0.043 (0.155) | 0.042 (0.155) |
| factor(country)China | −0.307** (0.151) | −0.305** (0.151) |
| factor(country)France | −0.179 (0.149) | −0.177 (0.148) |
| factor(country)Germany | 0.045 (0.148) | 0.048 (0.148) |
| factor(country)Greece | −0.390 (0.249) | −0.402 (0.249) |
| factor(country)Hong Kong | −0.434** (0.203) | −0.422** (0.203) |
| factor(country)Hungary | −0.377 (0.248) | −0.378 (0.248) |
| factor(country)India | −0.344** (0.159) | −0.338** (0.159) |
| factor(country)Ireland | 0.211 (0.164) | 0.213 (0.164) |
| factor(country)Israel | −0.433* (0.248) | −0.438* (0.248) |
| factor(country)Italy | −0.052 (0.152) | −0.050 (0.152) |
| factor(country)Japan | −0.113 (0.148) | −0.106 (0.147) |
| factor(country)Mexico | 0.022 (0.186) | 0.028 (0.187) |
| factor(country)Netherlands | −0.142 (0.159) | −0.144 (0.159) |
| factor(country)Norway | 0.438** (0.185) | 0.447** (0.185) |
| factor(country)Poland | 0.192 (0.151) | 0.194 (0.151) |
| factor(country)Portugal | 0.218 (0.176) | 0.219 (0.176) |
| factor(country)South Africa | −0.209 (0.249) | −0.213 (0.249) |
| factor(country)South Korea | −0.018 (0.166) | −0.011 (0.166) |
| factor(country)Spain | −0.113 (0.170) | −0.109 (0.170) |
| factor(country)Sweden | 0.050 (0.153) | 0.051 (0.153) |
| factor(country)Switzerland | −0.077 (0.154) | −0.072 (0.154) |
| factor(country)Taiwan | −0.322* (0.176) | −0.318* (0.176) |
| factor(country)Thailand | 0.517** (0.249) | 0.512** (0.250) |
| factor(country)Turkey | −0.385 (0.249) | −0.371 (0.249) |
| factor(country)United Kingdom | 0.114 (0.148) | 0.115 (0.148) |
| factor(country)United States | 0.024 (0.145) | 0.026 (0.145) |
| factor(country)United States of America | −0.415* (0.250) | −0.406 (0.250) |
| Constant | 0.888*** (0.175) | 0.912*** (0.171) |
| Observations | 635 | 635 |
| $R^2$ | 0.345 | 0.345 |
| Adjusted $R^2$ | 0.291 | 0.291 |
| Residual Std. Error (df = 585) | 0.202 | 0.202 |
| F Statistic (df = 49; 585) | 6.300*** | 6.302*** |

## 14.5 Firm size

I include specific tests on predicted withdrawal scale by firm size; while mid cap and small cap firms are more likely to withdraw than large cap firms, the result is not statistically significant. As the findings from the main empirics suggest; geographical revenue plays a larger role in predicting company withdrawals.



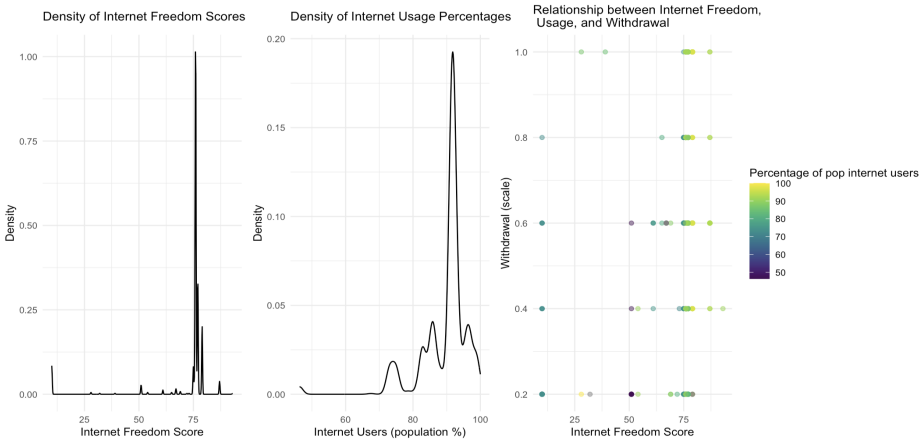

*add

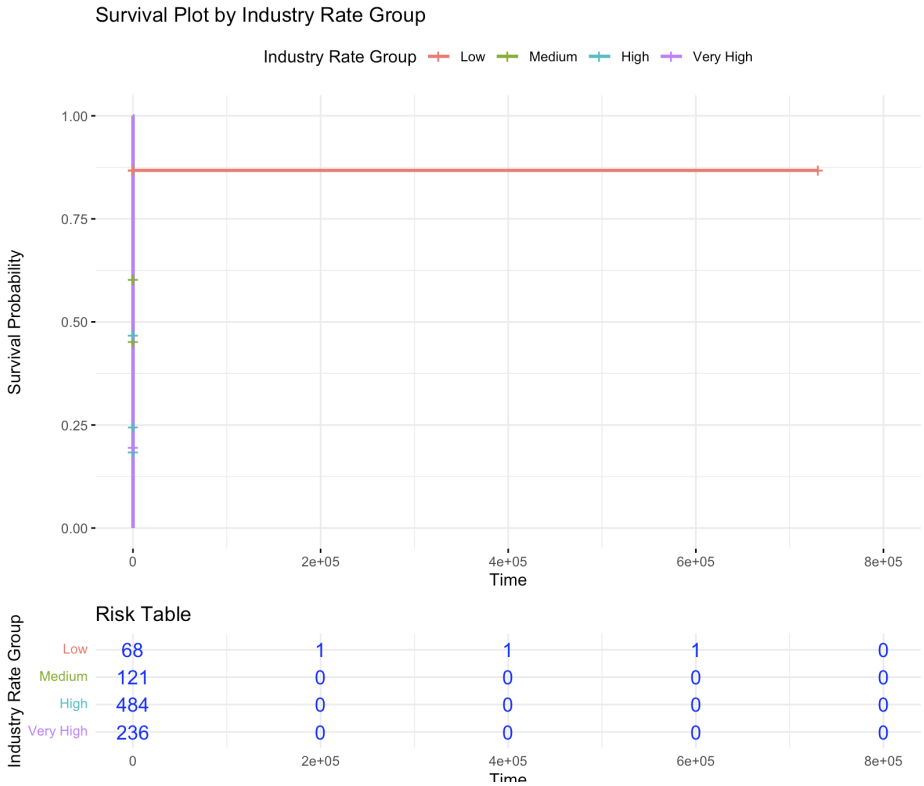
**Establishing the audience: Consumer pressure, boycotts and Online Pressure**

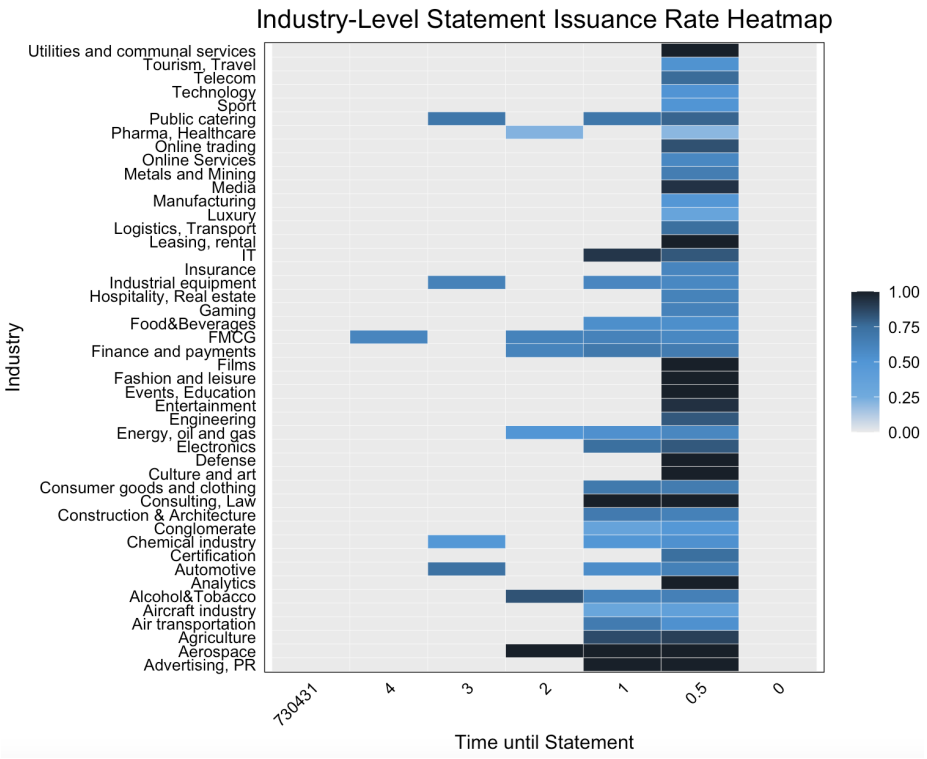
In order to make the argument that reputational concerns generate by a consumer audience influenced firm behaviour, I first demonstrate a relationship between consumer pressure and firm withdrawals. As per Guriev, Melnikov, and Zhuravskaya 2021; Lu, Huang, and S. Li 2022; Chen and Yang 2019 who have previously shown that higher levels of internet freedom, higher internet speeds, and more internet subscriptions in a region create greater public social awareness and more unrest, I proxy for public and media pressure through internet freedom levels (Freedom House 2022), and internet subscriptions (World Bank 2022). I

use univariate logistic regression with country fixed-effects. With controls, I find that MNCs headquartered in states with higher levels of media exposure and greater social awareness are more likely to withdraw from Russia. The results are depicted in the plots below.

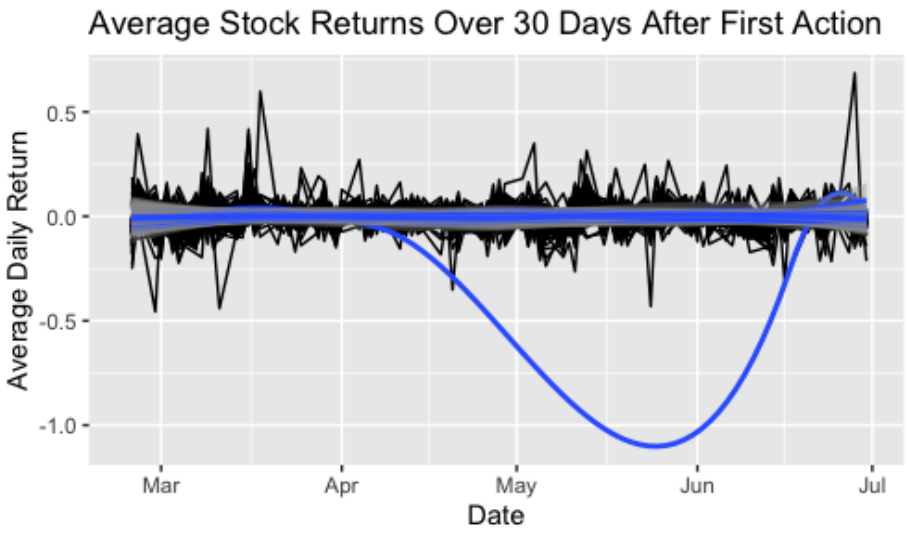



Survival Plot by Industry Rate Group
Industry Rate Group
Low
Medium
High
Very High
Survival Probability
1.00
0.75
0.50
0.25
0.00
0
2e+05
4e+05
6e+05
8e+05
Time
Risk Table
Industry Rate Group
Low 68 1 1 1 0
Medium 121 0 0 0 0
High 484 0 0 0 0
Very High 236 0 0 0 0
0
2e+05
4e+05
6e+05
8e+05
Time




## 14.6 Stock-market Returns



# 15 Robustness Checks

## 15.1 Cox Proportional Hazard Assumption Violation Checks

In the context of Cox proportional hazards models, the assumption is that the hazard ratios are constant over time. Scaled Schoenfeld residuals are a diagnostic tool used to test this assumption. These residuals are a transformation of the Schoenfeld residuals, which are used to assess whether the effect of a covariate changes over time. Essentially, if the proportional hazards assumption holds, these residuals should show no systematic pattern or trend when plotted against time.

I checked the proportional hazards assumption with scaled schoenfeld residuals using the cox.zph function from the `survival` package. This function performs a test of the proportional hazards assumption for a Cox regression model fit. The output below shows the generalized variance inflation factors (GVIFs) for each of the predictors in the model. The GVIF values are a generalization of the classic Variance Inflation Factor (VIF) for situations where predictors have more than one degree of freedom. These values give an indication of the presence of multicollinearity. This is the adjusted GVIF value, raised to the power of $1/(2Df)$. This adjustment allows for a better comparison across predictors with different degrees of freedom. A GVIF (or adjusted GVIF) value close to 1 suggests that the predictor is not highly collinear with the other predictors. Larger GVIF values indicate greater multicollinearity. Although there's no strict threshold, values above 5 (or sometimes 10) are often considered indicative of concerning multicollinearity. `industry rate`, `ungc member`, and `ESG Raw` all have GVIF values around 1, suggesting lower multicollinearity, while `firm size category` has a slightly higher value, but it's still within acceptable limits. I omit fixed effect variables via stratification in the Cox model because

they are not part of the proportional hazards function itself but rather modify the baseline hazard function. Thus, their inclusion in the VIF analysis is not informative.

| Variable | GVIF | Df | $GVIF^{(1/(2*Df))}$ |
|---|---|---|---|
| industry_rate | 1.090448 | 1 | 1.044245 |
| firm_size_category | 1.408081 | 2 | 1.089324 |
| ungc_member | 1.170677 | 1 | 1.081978 |
| esg_raw | 1.476131 | 1 | 1.214961 |

**Table 11.** Variance Inflation Factor (VIF) Analysis

The output of the cox.zph function in R provides a test of the proportional hazards assumption for each covariate in your Cox model, as well as a global test for the overall model. `industry rate` has a p-value of 0.92, and therefore there is no evidence to suggest a violation of the proportional hazards assumption, `ungc member` has a p-value of 0.54, it seems to meet the proportional hazards assumption, and `ESG Raw` the p-value is 0.16 indicating no significant evidence of violating the proportional hazards assumption, and `firm size category` has a p-value of 0.94, it also does not appear to violate the assumption. The GLOBAL row represents a test for the overall model. With a p-value of 0.44, it suggests that, taken as a whole, the model does not violate the proportional hazards assumption. Based on these results, there isn't strong evidence to suggest that any of the individual covariates violate the proportional hazards assumption.

| | coef | exp(coef) | se(coef) | z | p |
|---|---|---|---|---|---|
| industry_rate | 3.75 | 42.69 | 0.40 | 9.39 | 0.00 |
| firm_size_categoryMid Cap | 0.85 | 2.33 | 0.23 | 3.68 | 0.00 |
| firm_size_categorySmall Cap | 0.39 | 1.48 | 0.40 | 0.96 | 0.34 |
| ungc_member | -0.12 | 0.88 | 0.22 | -0.57 | 0.57 |
| esg_raw | -0.01 | 0.99 | 0.01 | -0.87 | 0.38 |

For the other models used, missing data is an issue, however in testing for multicolinearity, these VIF values are very close to 1, which is considered very low. This suggests that there is very little collinearity among these predictors. Low VIF values indicate that these variables are largely independent of each other in terms of their contribution to the model. For the test of the proportional hazards assumption, the p-value of 0.057 is slightly above 0.05, indicating that globally the model does not show strong evidence against the proportional hazards assumption, although it's close to the threshold. With a p-value of 0.053, it is marginally significant, however this is largely because the hazard (risk of withdrawal) is not constant over time, industry has a different impact on the risk of withdrawal in the early phase after a firm enters the study compared to later. Given my argument, and later split of the data between early and late withdrawals, this is not a large concern.

| Variable | VIF |
|---|---|
| industry_rate | 1.002410 |
| imputed_share_revenue_ru_2021 | 1.002960 |
| final_firm_size | 1.006665 |
| ungc_member | 1.006071 |

**Table 12.** Variance Inflation Factor (VIF) Results

| Variable | Chisq | df | p-value |
|---|---|---|---|
| industry_rate | 3.743 | 1 | 0.053 |
| imputed_share_revenue_ru_2021 | 1.227 | 1 | 0.268 |
| final_firm_size | 3.430 | 1 | 0.064 |
| ungc_member | 0.509 | 1 | 0.476 |
| GLOBAL | 9.185 | 4 | 0.057 |

**Table 13.** Cox Proportional Hazards Model Test for Proportional Hazards

### 15.2 Additional Controls

Company statements from both remaining and withdrawing firms revealed a number of stated reasons justifying why companies chose to remain, or why withdrawal would be a lengthy business. I therefore include a battery of additional control variables included from the KSE database. I add percentage share of company revenue in Russia as of 2021. I chose 2021 instead of 2022 since this was a more reliable measure of the company's revenue in Russia as a baseline. I included global revenue (2021) as a measure, in addition to firm size (company market cap), and assets in Russia (2022). I chose assets in Russia in 2022 over 2021 since this was cited multiple times in firm statements as a key concern by companies that if they withdrew their assets would be seized by the government in retaliation. It was also used to explain late withdrawals. I also included allied state headquarters (if the company was headquartered in a state allied with Russia). Since a number of companies also cited concern for remaining staff members in Russia and protecting staff families as a reason not to leave Russia, I also controlled for number of staff for companies that had disclosed this information. I included ESG scores for companies that disclosed this information, and UNGC membership. The cox model below depicts the cox proportional hazard model output with (1) a bivariate regression of industry withdrawal rate on likelihood of withdrawing quickly, model (2) includes all controls without fixed-effects, and model (3) shows all controls plus two-way fixed effects at country and industry levels to control for unobserved heterogeneity across countries and industries that might be driving the effect. I clustered standard errors in model 3 at industry level. The results remain robust for industry withdrawal, with the effect dropping only slightly when adding controls and fixed-effects. As suspected, having assets in Russia,a greater share of revenue in Russia, and more staff members, slows down a company's withdrawal rate. Larger firms are less

likely to withdraw, and ESG has a positive but statistically insignificant effect on withdrawals. UNGC membership is associated with slower withdrawals, but this effect disappears after adding fixed-effects. One variable I do not include here is years in Russia - I could not find this information for a large proportion of the sample here and therefore chose to omit it. It is not greatly concerning since percentage of revenue, and number of locations in Russia, likely overlap with time in Russia.

**Table 14.** Cox Proportional Hazard Models: Sector Network Effect on Withdrawal with country-industry fixed effects

| | Survival Risk | | |
|---|---|---|---|
| | Rate of Withdrawal (Withdrawal Dummy) | | |
| | (1) | (2) | (3) |
| Sector Network Effect | 3.246*** | 2.851*** | 2.735*** |
| | (0.048) | (0.123) | (0.292) |
| Percentage Share of Company Revenue in Russia (2021) | | 0.037*** | −0.051 |
| | | (0.008) | (0.016) |
| Global Revenue 2021 | | −0.00000*** | −0.00001** |
| | | (0.00000) | (0.00000) |
| Assets in Russia (2022) | | −0.00002 | −0.0002 |
| | | (0.0003) | (0.0004) |
| Allied State HQ | | −15.710 | −16.520*** |
| | | (559.207) | (3,780.943) |
| Number of Staff (2021) | | 0.001*** | 0.002*** |
| | | (0.0001) | (0.0002) |
| Final Firm Size | | −0.021 | −0.028 |
| | | (0.058) | (0.059) |
| ESG Score | | 0.004 | 0.006 |
| | | (0.003) | (0.003) |
| UNGC Member | | −0.180* | −0.048 |
| | | (0.095) | (0.101) |
| FE (Country, Industry) | No | No | Yes |
| Observations | 20,885 | 2,608 | 2,608 |
| $R^2$ | 0.215 | 0.314 | 0.541 |
| Max. Possible $R^2$ | 1.000 | 0.999 | 0.999 |
| Log Likelihood | −85,770.650 | −8,743.213 | −8,218.191 |
| Wald Test | 4,553.900*** (df = 1) | 547.630*** (df = 9) | 3,265,820.000*** (df = 64) |
| LR Test | 5,062.328*** (df = 1) | 983.070*** (df = 9) | 2,033.115*** (df = 64) |
| Score (Logrank) Test | 4,818.449*** (df = 1) | 1,055.638*** (df = 9) | 2,072.040*** (df = 64) |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 |

**Table 15.** Cox Proportional Hazards Model Summary

| Variable | Coefficient | Exp(Coefficient) | SE(Coefficient) | Z | P |
|---|---|---|---|---|---|
| Industry Rate | 3.192e+00 | 2.434e+01 | 3.781e-01 | 8.443 | < 2e-16*** |
| Firm size (Comp Market Cap) | 6.416e-01 | 1.900e+00 | 5.488e-01 | 1.169 | 0.242335 |
| Industry: Utilities and communal services | 8.812e-01 | 2.414e+00 | 1.116e+00 | 0.790 | 0.429661 |
| factor(Country of HQ)Austria | -7.665e-01 | 4.646e-01 | 7.464e-01 | -1.027 | 0.304414 |
| factor(Country of HQ)Belgium | 2.978e-01 | 1.347e+00 | 7.597e-01 | 0.392 | 0.695013 |
| factor(Country of HQ)Brazil | 3.139e-02 | 1.032e+00 | 1.099e+00 | 0.029 | 0.977221 |
| factor(Country of HQ)USA | 2.679e-01 | 1.307e+00 | 4.708e-01 | 0.569 | 0.569369 |
| GeoRev | -4.434e-02 | 9.566e-01 | 4.699e-02 | -0.944 | 0.345287 |
| Likelihood ratio test=387 on 87 df, p < 2.2e-16<br>n=909, number of events=550 | | | | | |

## 15.3 Social Media: scale

**Table 16.** Probit Models: Effect of Firm Exit Decisions on Issuing Social Media Statements (Scales)

| | Issues Social Media Statement | |
|---|---|---|
| | Statement | |
| | (1) | (2) |
| Scale of Withdrawal | 0.305*** | 0.006 |
| | (0.052) | (0.068) |
| UNGC member | | 0.058* |
| | | (0.034) |
| Big Firm Dummy | | 0.071* |
| | | (0.041) |
| ESG score | | 0.002** |
| | | (0.001) |
| Allied State HQ | | −0.352 |
| | | (0.283) |
| Constant | 0.295*** | 0.221 |
| | (0.037) | (0.268) |
| FE (Country, Industry) | No | Yes |
| Observations | 1,280 | 950 |
| Log Likelihood | −912.832 | −535.752 |
| Akaike Inf. Crit. | 1,829.664 | 1,173.504 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 | |

**Table 17.** Probit Models: Effect of Early Firm Exit Decisions on Issuing Social Media Statements (Scales)

| | Issues Social Media Statement | |
|---|---|---|
| | Statement | |
| | (1) | (2) |
| Scale of Withdrawal | 0.285*** | −0.022 |
| | (0.055) | (0.075) |
| UNGC member | | 0.057* |
| | | (0.034) |
| Big Firm Dummy | | 0.072* |
| | | (0.041) |
| ESG score | | 0.002** |
| | | (0.001) |
| Allied State HQ | | −0.366 |
| | | (0.284) |
| Constant | 0.330*** | 0.240 |
| | (0.035) | (0.269) |
| FE (Country, Industry) | No | Yes |
| Observations | 1,279 | 950 |
| Log Likelihood | −916.042 | −535.709 |
| Akaike Inf. Crit. | 1,836.084 | 1,173.418 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 | |

### 15.4 Donations

I collected data on whether firms specifically referenced providing aid, humanitarian relief, or donations to Ukraine in the statement they released, either through their website or through social media. Of my sample of firms, 885 firms included information on donations and humanitarian aid provisions in their statements, while 383 did not. Below I consider the effect of withdrawing on explicitly declaring donations, and whether companies that issued social media statements were more likely to declare donations. While the expectation would be that withdrawing firms would be more likely to draw attention to their efforts for social capital gains, it is also possible that remaining firms would be more likely to focus on donations and humanitarian aid in an effort to divert attention from their decision to remain in Russia. In table 6 I use a mixture of linear regression and probit models to evaluate the effect of withdrawing firms and statements that cite humanitarian aid/donations to Ukraine. Table 6 shows that withdrawing firms are more likely to explicitly mention humanitarian aid and donations, using a scale of withdrawal (a weighted variable between 0-1 between 'digging in', 'buying time', 'suspending operations', 'scaling back', and 'withdrawal'). Table 7 uses a withdrawal dummy, and table 8 examines early withdrawing firms.

**Table 18.** OLS Models with Two-Way FE: Grade of Firm Withdrawl on Donations/Humanitarian Aid

| | Explicit Humanitarian Aid/Donations to Ukraine | |
|---|---|---|
| | Statement | |
| | (1) | (2) |
| Scale of Withdrawal | 0.537*** | 0.355*** |
| | (0.046) | (0.055) |
| UNGC member | | 0.020 |
| | | (0.028) |
| Big Firm Dummy | | 0.102*** |
| | | (0.034) |
| ESG score | | 0.006*** |
| | | (0.001) |
| Allied State HQ | | −0.154 |
| | | (0.230) |
| Constant | 0.346*** | −0.277 |
| | (0.033) | (0.217) |
| FE (Country, Industry) | No | Yes |
| Observations | 1,268 | 946 |
| Log Likelihood | −748.511 | −335.718 |
| Akaike Inf. Crit. | 1,501.021 | 773.436 |
| *Note:* | | *p<0.1; **p<0.05; ***p<0.01 |

**Table 19.** Probit Models: Exiting Firms and Donations/Humanitarian Aid

| | Explicit Humanitarian Aid/Donations to Ukraine | |
|---|---|---|
| | | Statement |
| | (1) | (2) |
| Withdrawal Dummy | 0.147*** | 0.096*** |
| | (0.029) | (0.030) |
| UNGC member | | 0.013 |
| | | (0.028) |
| Big Firm Dummy | | 0.103*** |
| | | (0.034) |
| ESG score | | 0.006*** |
| | | (0.001) |
| Allied State HQ | | −0.290 |
| | | (0.232) |
| Constant | 0.658*** | −0.072 |
| | (0.015) | (0.218) |
| FE (Country, Industry) | No | Yes |
| Observations | 1,268 | 946 |
| Log Likelihood | −800.255 | −351.539 |
| Akaike Inf. Crit. | 1,604.509 | 805.078 |
| *Note:* | | *p<0.1; **p<0.05; ***p<0.01 |

**Table 20.** Probit Models: Early Exiting Firms on Donations/Humanitarian Aid

| | Explicit Humanitarian Aid/Donations to Ukraine |
|---|---|
| | Statement |
| Withdrawal Dummy (Early) | 0.109*** |
| | (0.042) |
| UNGC member | 0.011 |
| | (0.028) |
| Big Firm Dummy | 0.102*** |
| | (0.034) |
| ESG score | 0.007*** |
| | (0.001) |
| Allied State HQ | −0.283 |
| | (0.233) |
| Constant | −0.095 |
| | (0.219) |
| FE (Country, Industry) | No |
| Observations | 946 |
| Log Likelihood | −353.286 |
| Akaike Inf. Crit. | 808.571 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

## 16 Text Analysis

I collected 300 statements issued by companies from my sample.

### 16.1 Sample of Texts

Table 21: Sample of Company Statements on Russian-Ukranian War

| Company | Statement Snapshot | Withdraw Dummy |
|---|---|---|
| earthdaily analytics corp | EarthDaily Analytics Corp. ("EDA" or the "Company"), a vertically-integrated data processing and ana | 1 |
| netscout systems inc | All of us at NETSCOUT are closely following the tragic events unfolding in Europe after the Russian | 1 |
| acerinox sa | Acerinox ceased its commercial activity in Russia as soon as the invasion in Ukraine began and has a | 0 |
| davide campari milano nv | Campari Group is present in Russia since 2011 and currently employs 122 Camparistas in the Country. | 0 |
| novartis ag | ...The global Novartis Emergency Management (NEM) team is engaged closely with the Ukraine Country N | 0 |
| volvo ab | | 0 |

Continued on next page

**Table 21 – continued from previous page**

| Company | Statement Snapshot | Withdraw Dummy |
|---|---|---|
| jetbrains sro | As a response to the Russian Federation's invasion of Ukraine, we announced that we would be suspend | 1 |
| ingka holding bv | Many thanks for your continuous work and commitment to Human Rights. The war in Ukraine is a human t | 0 |
| tchibo manufacturing gmbh & co kg | Tchibo had been operating in Russia since 1994, selling coffee products in retail outlets in Russia | 0 |
| rakuten group inc | As a member of the international community, Rakuten sincerely hopes for a peaceful resolution to thi | 0 |
| valio oy | "We strictly condemn Russia's attack on independent Ukraine. Ethically, Valio cannot continue operat | 1 |
| polpharma | In connection with the ongoing war on the territory of Ukraine, we would like to express our deepest | 0 |
| hp inc | | 0 |
| boston consulting group inc | It is horrifying to witness the impact of the escalating war in Ukraine on millions of people. Yeste | 1 |

Continued on next page

**Table 21 – continued from previous page**

| Company | Statement Snapshot | Withdraw Dummy |
|---|---|---|
| morgan advanced materials plc | We fully support the sanctions that have been put in place against Russia and we have ceased all tra | 1 |
| intesa sanpaolo spa | I'm pleased to respond to your kind request. Intesa Sanpaolo has subsidiaries both in Russia (Banca | 0 |
| vitol | Just over a decade. Until recently we were invested in Ukrainian gas production. This was divested | |
| boiron sa | In Russia, Boiron froze its investments, halted advertising campaigns, canceled new product launches | 0 |
| yokohama rubber co ltd | As the Yokohama Rubber Group's business activities expand globally, we believe in the importance of | 0 |
| mckinsey & company inc | Since the Russian government invaded Ukraine, we have focused first and foremost on ensuring the saf | 0 |
| abbvie inc | 15 April 2022 ...Hope all is well and apologies for the delay in getting back to you. For your quest | 0 |
| fm global | As we continue to watch the unfolding conflict, FM Global has made the decision to stop renewing any | 0 |

Continued on next page

**Table 21 – continued from previous page**

| Company | Statement Snapshot | Withdraw Dummy |
|---|---|---|
| baker tilly us llp | On 16 March 2022, Baker Tilly International announced we will separate member firms in Russia and Be | 1 |
| general electric | GE Healthcare and the GE Foundation have donated $4.5M in medical equipment and humanitarian aid to | |
| teradata corp | eradata stopped conducting business in Russia earlier this month, and has ceased customer interactio | 0 |
| yum! brands inc | Yum! Brands is suspending operations of KFC company-owned restaurants in Russia and finalizing an ag | 0 |
| orlen lietuva ab | Since the beginning of the war in Ukraine, PKN ORLEN, like other European counterparties, has not ma | 1 |
| compagnie de saint gobain sa | Saint-Gobain reiterates its support and solidarity with Ukraine, its population and all its Ukrainia | 0 |
| chevron corp | We have no operations or investments in Ukraine. We recently donated $5 million to the World Food Pr | 0 |
| airbnb inc | Key Takeaways Airbnb has taken a number of steps to support our Host and guest community Throughout | 0 |

Continued on next page

**Table 21 – continued from previous page**

| Company | Statement Snapshot | Withdraw Dummy |
|---|---|---|
| e.on | [Initial response from E.ON] thanks for your request. E.ON does not have any operating business in U | |
| kesko oyj | Kesko has decided to discontinue the purchase of Russian products and the export of food products to | 1 |
| binance | A hot topic now, but there is much misunderstanding. Very few people have a deep understanding of al | |
| wintershall dea ag | Wintershall Dea is active in Russia since 30 years and is today involved in three onshore gas produc | 0 |
| allianz se | ...Thank you for the invitation. For the time being, Allianz is unable to respond to the survey, but | 0 |
| stora enso oyj | We are shocked and saddened by Russia's attack on Ukraine and its impact on international security a | 0 |
| kimberly-clark corp | First and foremost, Kimberly-Clark is deeply saddened by the war in Ukraine and our hearts go out to | 0 |
| glencore plc | Glencore condemns the actions taken by the Russian government against the people of Ukraine. We have | 0 |

Continued on next page

**Table 21 – continued from previous page**

| Company | Statement Snapshot | Withdraw Dummy |
|---|---|---|
| deutsche bank ag | We are deeply concerned about the attack on a sovereign European country and that the borders in Eur | 0 |
| emerson electric co | We are deeply concerned about the Russia military actions and the humanitarian crisis in Ukraine. O | 0 |
| eli lilly and co | A Year Later, Support Continues Over the last year, Lilly, our affiliates, and 38,000 employees worl | 0 |
| zoom video communications inc | | 0 |
| sumitomo electric industries ltd | Since the onset of this crisis on February 24th, Sumitomo Corporation has put the utmost importance | 0 |
| msc industrial direct co inc | We are profoundly shocked and deeply concerned by the Russian invasion of Ukraine and the ongoing an | 0 |
| kuehne und nagel international ag | On March 11, 2022, Kuehne+Nagel announced that it pledges emergency aid to Ukraine in the amount of | 0 |
| diesel | Dear Business & Human Rights Resource Centre, Thank you for reaching out and sharing your survey. Wi | |

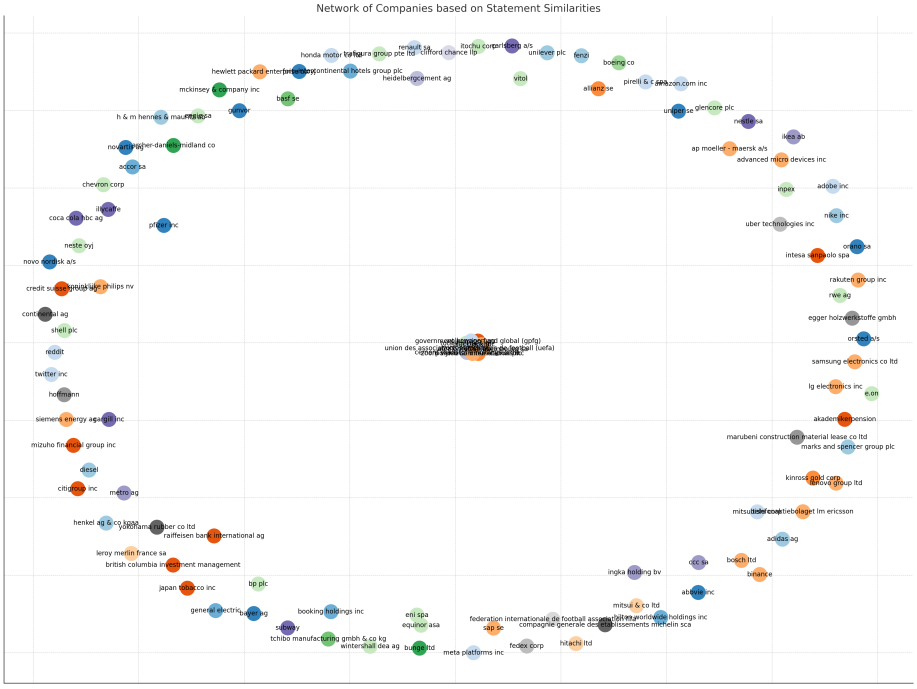
Continued on next page

**Table 21 – continued from previous page**

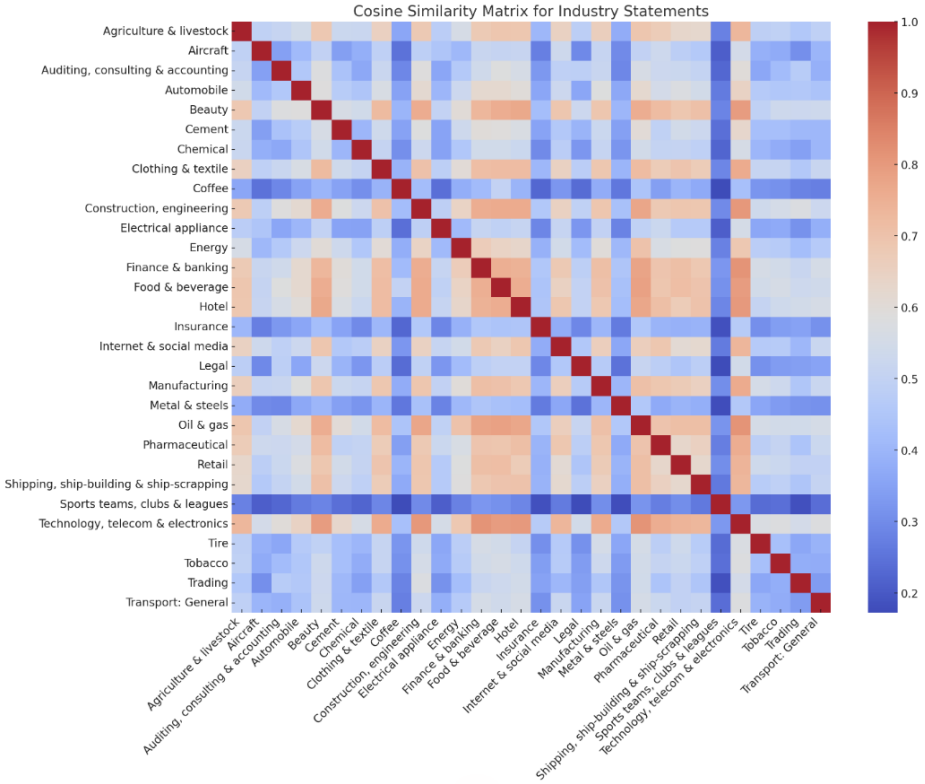
| Company | Statement Snapshot | Withdraw Dummy |
|---|---|---|
| barilla holding spa | We are constantly monitoring the evolution of the war in Ukraine and we are deeply shocked by the te | 0 |
| rwe ag | Many thanks for inviting us to respond to your questions regarding business activities in Ukraine an | |
| ernst & young llp | EY can confirm that the separation of the Russian and Belarus EY member firms from the EY global org | 1 |
| par pacific holdings inc | The owner of Hawaii's oil refinery said Thursday that island consumers should not expect oil supply | 0 |

## 16.2 Network Plots

The graph represents a network plot of a sub-sample of firm statements from the data frame. Each node represents a company. The edges (connections) between nodes indicate a high cosine similarity in their statements. Nodes are color-coded based on the sector they belong to. Companies from the same sector have the same color. This visualization provides an overview of how companies are connected based on the similarity of their statements, with the added context of their sub-sectors.



## 16.3 Cosine Similarity



### 16.4 Word-Embedding

The word frequency and topic models used in this paper reveal relationships between the frequency of referrals to multi-national stakeholder initiatives and UN-adjacent institutions and the statements issued by remaining firms. These methods also show how remaining firms appeal to justify their continued presence in Russia, and the type of language they avoid. To further analyze explicit differences in the language used by withdrawing and remaining firms in the statements released by these firms, I utilise word-embeddings. Word vectors establish parameter estimates of artificial neural networks that are then used to predict the occurrence of a word within a sequence of text (Rheault and Cochrane 2020). These models can help capture and illustrate the relationship between how remaining and withdrawing firms refer to 'Russia' and 'Ukraine'. Where topic-modelling and word-frequency treats words in isolation, word-embedding allows for n-grams to consider how these terms are used in a surrounding context.

I split the dataset up into three corpus; one corpus containing all firm statements, one corpus containing firms that are explicitly 'digging in' or 'buying time' (remaining firms), and one corpus that is explicitly 'withdrawing' and 'scaling back' (withdrawing firms). I chose to omit 'suspension' firms (with 'C' grades) since these firms may be merely complying with sanctions. These thresholds therefore compare 'failing grade firms' (i.e. D's and F's) with 'passing grade' firms (A's' and B's).

To ensure randomness and mitigate potential sequence bias in the text data, we first shuffled the statement column of the statements in the datasets using a fixed seed (set.seed(42L)) for reproducibility. Text preprocessing involved several steps to standardize and simplify the dataset for analysis. All text was converted

to lowercase to maintain consistency and avoid duplication of words based on case differences. All punctuation marks were removed (gsub("[[:punct:]]", "", text)) to focus solely on the words. The text was tokenized into individual words using a space tokenizer. This step breaks down the text into its basic units for analysis. Commonly used words (such as "the", "and", "in") that generally do not contribute to the semantic meaning in the analysis context were removed. This step was crucial to focus on more meaningful words in the subsequent analysis. A vocabulary was created from the preprocessed tokens, I then pruned this vocabulary to retain only those words that appeared more than a specified minimum count (COUNTMIN = 10), ensuring that the analysis focused on words with sufficient representation in the dataset. I constructed a Term Co-occurrence Matrix (TCM) considering the window size (WINDOWSIZE = 6) around each word. This matrix quantifies how frequently each pair of words occurs together within the defined context window, capturing the local context of words in the text. The Global Vectors for Word Representation (GloVe) model was employed to generate word embeddings. I initialized the GloVe model with a specified dimension (DIM = 300) and set the maximum number of co-occurrence (xmax = 100). The model was trained on the TCM for a defined number of iterations (ITERS = 100) to produce word vectors. These vectors encapsulate semantic relationships between words based on their co-occurrence patterns in the corpus. To visualize the high-dimensional word vectors, we applied Uniform Manifold Approximation and Projection (UMAP) for dimension reduction, transforming the word vectors into a 2D space. UMAP parameters were set to number of components = 2, metric = "cosine", number of neighbors = 25. The resulting 2D embeddings are shown in the figures below, which visualiizes the semantic relationships among words in each of the three corpus.

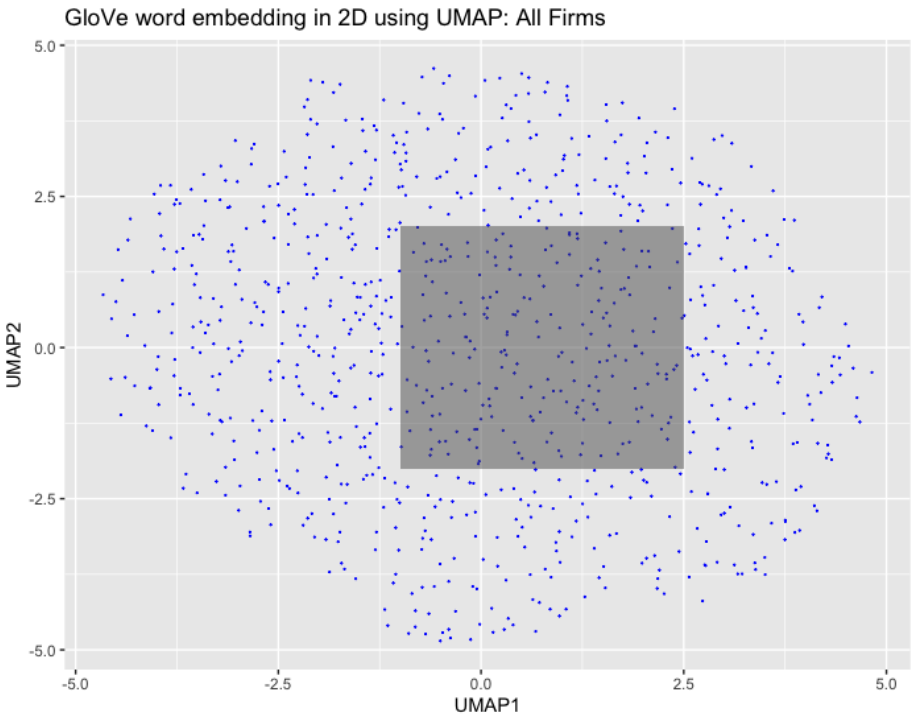



**Fig. 3.** UMAP dimensions: All Firm Statements

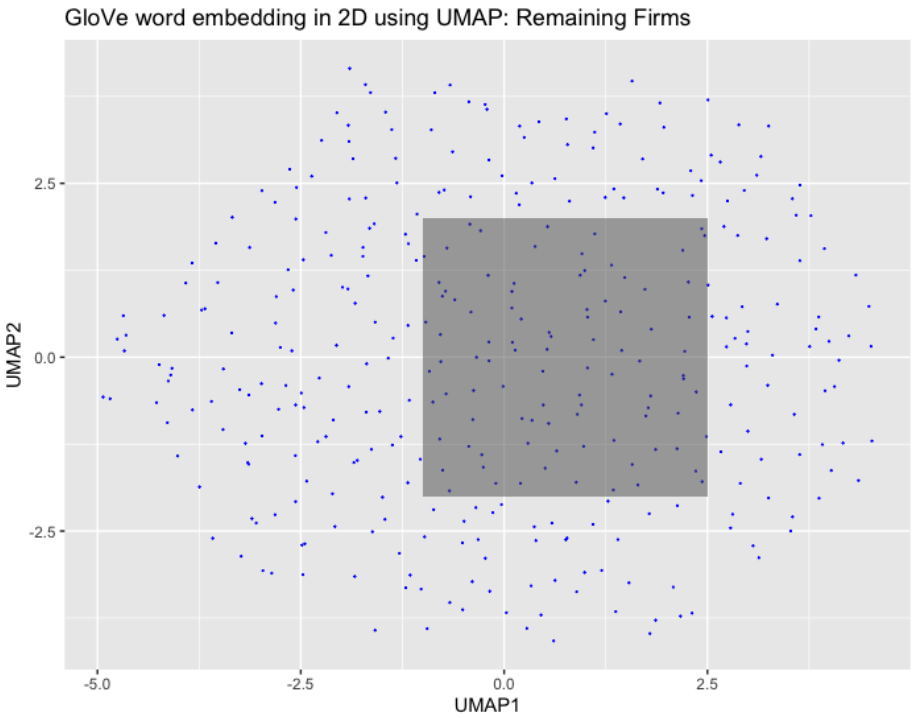



**Fig. 4.** UMAP dimensions: Remaining Firms Statements

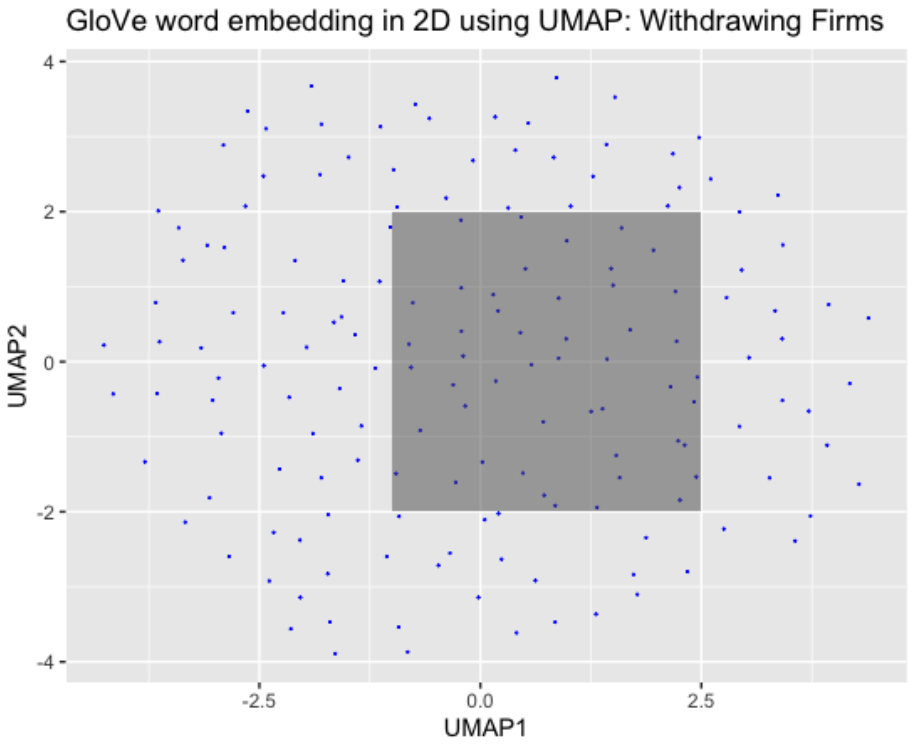



**Fig. 5.** UMAP dimensions: Withdrawing Firms Statements

I then focused on extracting the GloVe word embedding vector for the words "Ukraine" and "Russia" among the three corpuses. The embedding vector is a numeric representation capturing the semantic essence of these terms based on its usage and context in each dataset. I identify the top 25 words most closely related to "Ukraine" and "Russia" separately, based on their cosine similarity scores. The results below reveal interesting patterns.

**GloVe: Ukraine** In the GloVe model for withdrawing firms, the words closely associated with "Ukraine" appear to relate to the business and operational impacts of the withdrawal. Notable terms like "invasion," "conflict," "operations," and "war" suggest a narrative focused on the response to conflict and its implications for business activities. The presence of words like "employees," "office," and "well-being" indicates a concern for personnel and organizational health. The distribution of words across the UMAP dimensions shows a somewhat dispersed clustering, with "invasion" and "conflict" being more isolated, which could indicate a unique contextual usage separate from other operational terms.

For the remaining firms, the terms related to "Ukraine" emphasize the human and emotional aspects, with words like "safety," "children," "crisis," and "saddened" being prominent. This suggests a focus on the humanitarian aspect of the situation. The term "condemn" implies a stance or action taken by the firms, and the presence of "investment" and "temporary" might reflect the firms' strategic considerations during the conflict. The spread of words here is more concentrated in the center of the plot, indicating a possibly more unified context in which these terms are used across firms that remained and issued statements.

The major differences highlighted here are three-fold. First, in terms of thematic focus, withdrawing firms' corpus has a stronger emphasis on business and operational terms, whereas the remaining firms' corpus focuses more on the human and emotional impact of the situation in Ukraine. This finding adds weighs to the argument that remaining firms tended to redirect attention towards the situation in Ukraine as a 'humanitarian crisis', and divert attention away from continuing to operate within Russia, or any mention of Russia. Furthermore, in terms of differences in contextual usage between both word-embedding models,

specific terms like “invasion” and “war” in the withdrawing firms’ corpus have a distinct placement, potentially indicating a particular emphasis on the conflict aspect. In contrast, the remaining firms’ corpus features terms like “crisis” and “safety” more centrally, highlighting that these firms attempt to avoid using antagonizing language around the situation. In regards to the clustering of observations, the withdrawing firms’ words show more dispersion, which may indicate a wider variety of contexts, while the remaining firms’ words cluster more tightly, suggesting a more focused narrative. This again would reinforce the idea that remaining firms learned from each other’s strategies to help protect their reputation.

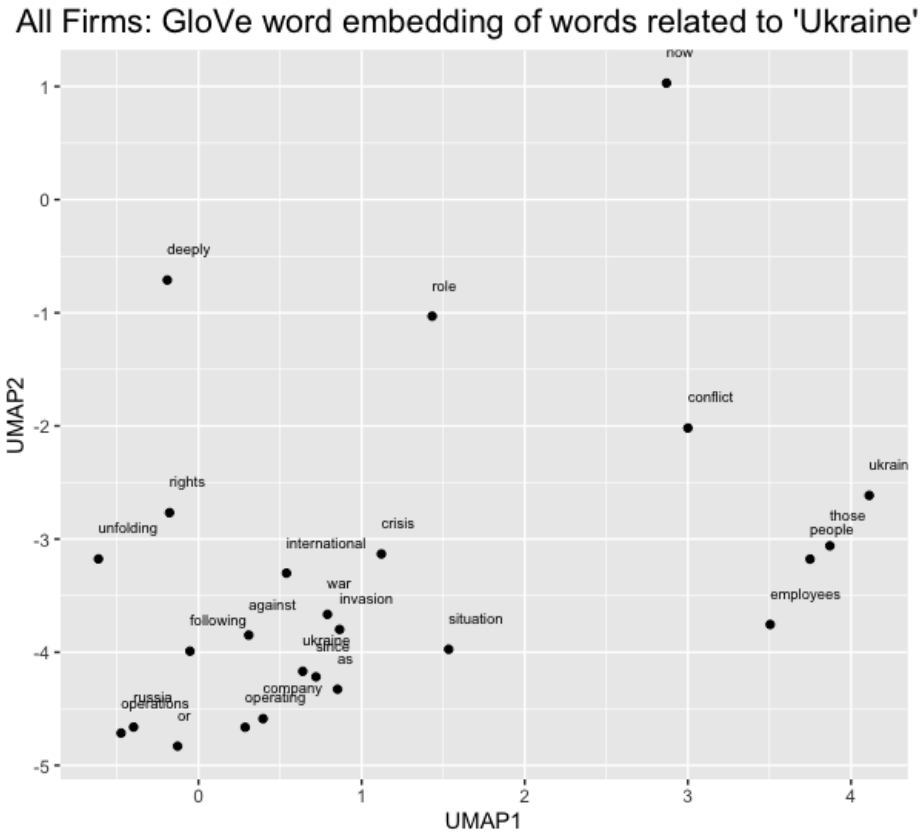



**Fig. 6.** GloVe Word Embedding: All Firms (Ukraine)

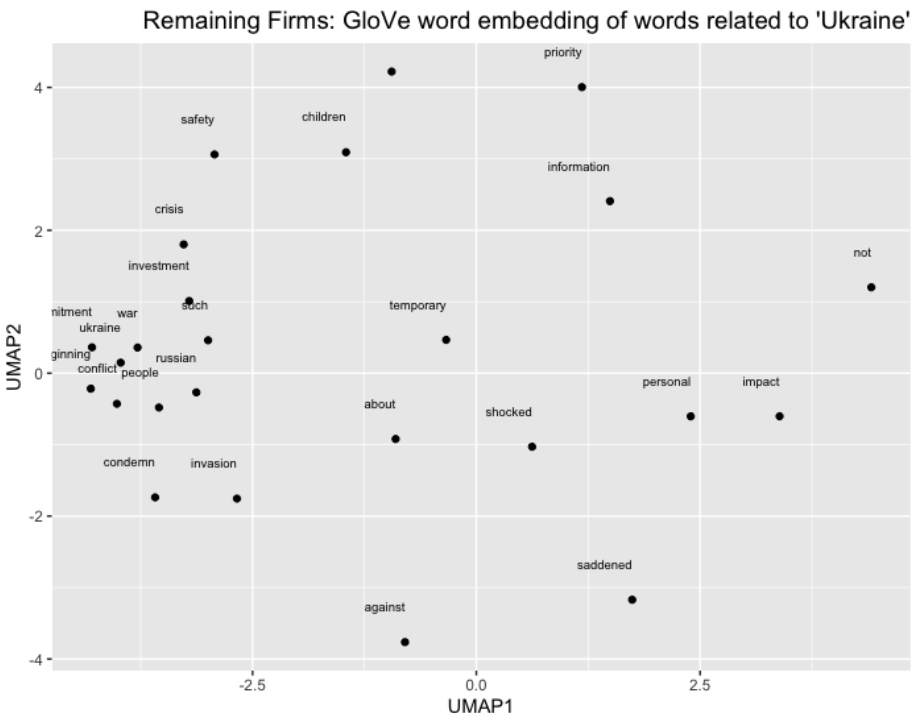



**Fig. 7.** GloVe Word Embedding: All Firms (Russia)

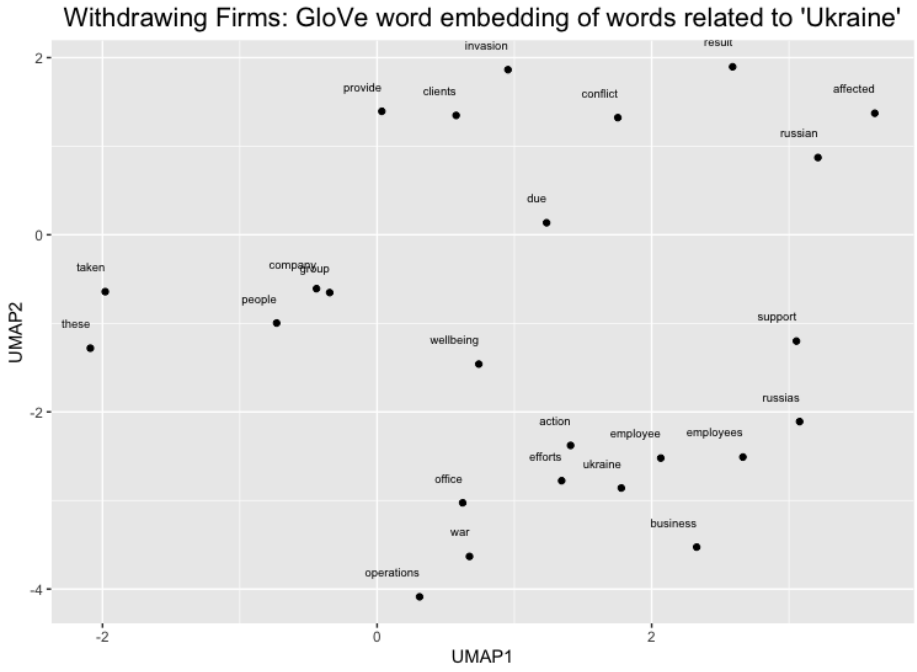



**Fig. 8.** GloVe Word Embedding: Withdrawing Firms (Ukraine)

**GloVe: Russia** In the GloVe models of withdrawing firms in relation to the embedded term 'Russia', there is again a similar strong emphasis on the business operations, with dominant terms such as "operations," "services," "sales," and "measures". This indicates a narrative centered on the logistical and practical implications of withdrawal. Unlike Ukraine, embedding around 'Russia' has a strong temporal dimension, with words like "longer," "today," "time," and "new" indicating considerations about the timing and duration of actions or policies related to Russia.

For remaining firms, the GloVe word embedding for 'Russia' in statements reveals an emphasis on continuity; words such as "continue," "suspended," and "ongoing" reveal a tension between maintaining ongoing operations and introducing changes or suspensions in response to the situation. Unlike the withdrawing corpus, the remaining corpus uses Russia in relation with peace, rather than condemnation. The contextual presence of "peace" in relation to Russia suggests that remaining firms attempt to shift audience attention towards a future state of peace, while avoiding referencing the current war. Remaining firms also highlight their financial embeddedness in Russia, with terms such as "decisions," "global," and "investment". However, as the results in this paper reveal, geographical exposure to Russia does not predict company decisions to withdraw, meaning these firms are attempting to emphasise their investment and asset risks to explain their decisions to remain or buy time. Similarly, these firms make multiple references to their obligations to their Russian employees as a reason to remain, with the presence of "colleagues," "employees," "group," and "organization" points to discussions around the impact on personnel and corporate structures.

Both corpora address the human element, but the withdrawing firms' corpus seems to highlight the personal impact of withdrawal decisions more, as evidenced by the term "families," while the remaining firms' corpus focuses more on the collective experience of employees and organizational dynamics. The withdrawing firms' corpus shows a more scattered clustering of terms, possibly reflecting a broader range of topics or contexts in discussions about Russia. In contrast, the remaining firms' corpus presents a tighter clustering, once again potentially indicating a more unified narrative suggestive of 'learning' between remaining firms.

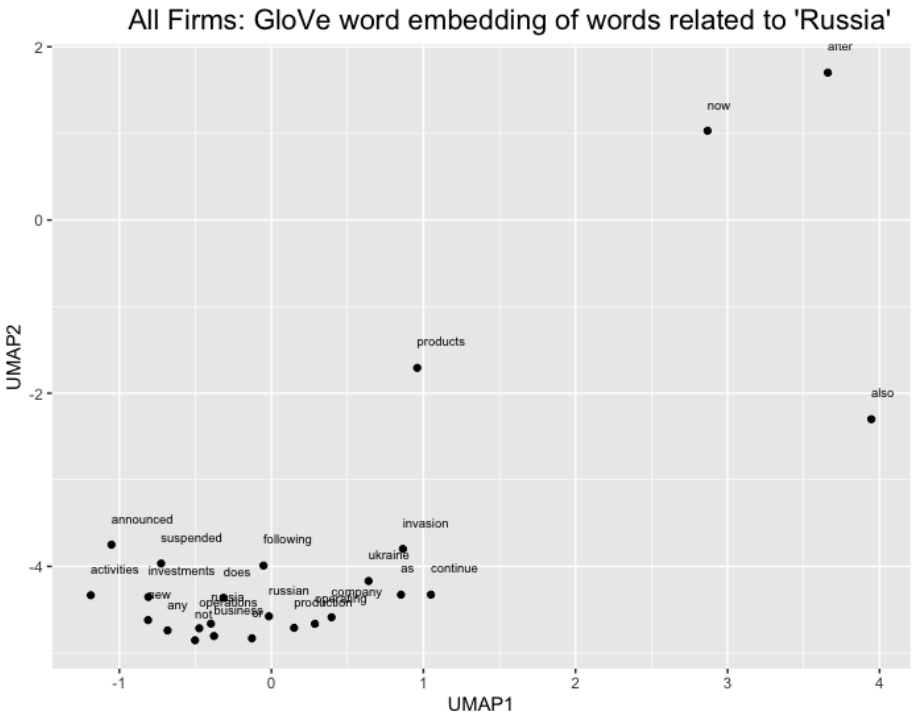



**Fig. 9.** GloVe Word Embedding: Withdrawing Firms (Russia)

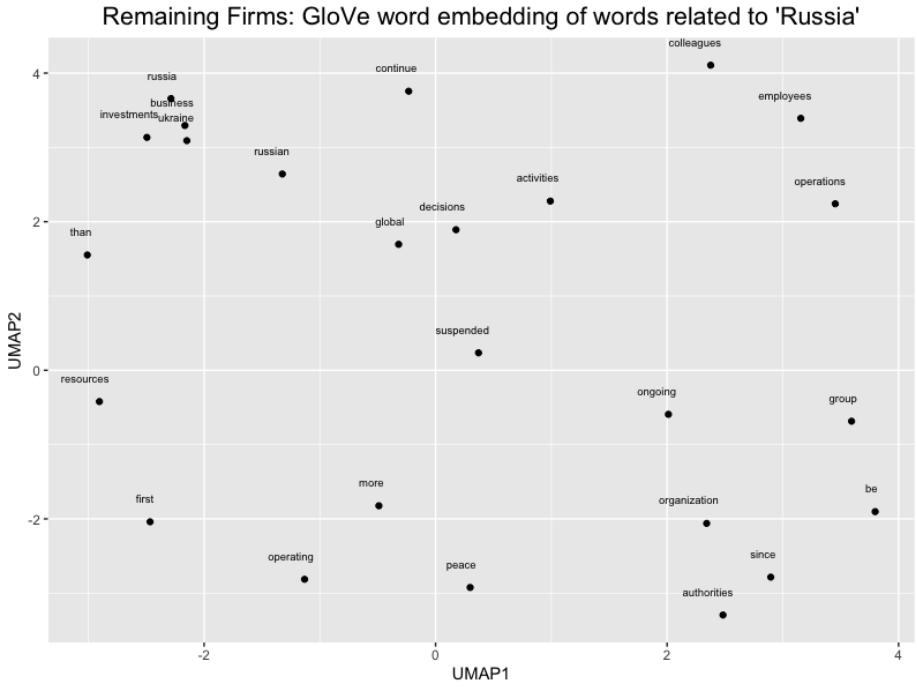



**Fig. 10.** GloVe Word Embedding: Remaining Firms (Russia)

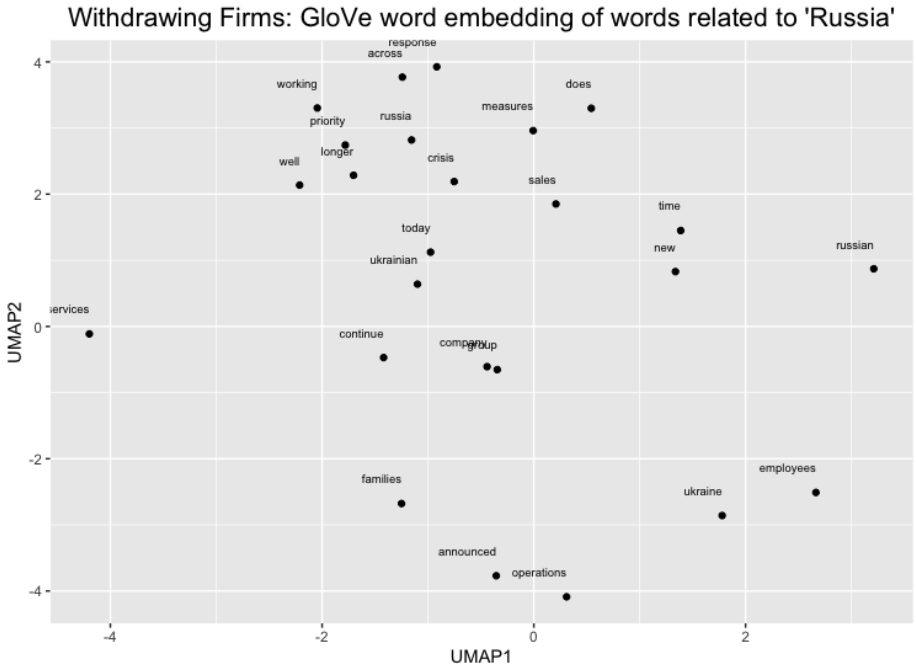



**Fig. 11.** GloVe Word Embedding: Remaining Firms (Ukraine)

## 16.5 Structural Topic Modelling

To further analyze the differences between withdrawing and remaining firm statements, I analyze these texts using structural topic modelling (M. E. Roberts, Stewart, Tingley, et al. 2014). STM is a mixed-membership style of topic modelling that uses LDA. Unlike simple LDA, it also enables the use of document-level covariates. In this analysis I focus on two elements; scale of withdrawal of firms (with five categories) and time (day statement was published). I use the *STM* package in *R*, that employs a variational expectation-maximizing algorithm (M. E. Roberts, Stewart, and Tingley 2016). I specify topic model prevalence as a function of the grading of the firm and a smooth function of date of posting. After pre-processing my text, I removed 2100 of 4090 terms (2100 of 28004 tokens) due to frequency, leaving the corpus with 253 documents, 1990 terms and 25904 tokens. I set the maximum iterations to 75, and the model ran for 29 iterations before converging. The relative changes were small, 1.788e-05, 1.442e-05, 1.337e-05 etc. In terms of notation, $K = 7$ was the userselected number of topics due to the smaller sample size of documents, $V = 1990$ was the size of the volcabulary in use, and $D = 253$ the size of documents. There are two relavent matrix parametres involved in the STM;  represents a row-normalized topics-by-vocabulary matrix of topic-word distributions. This represents the probability of observing the v-th word in topic n = k. In the second matrix;  should be read as the row-normalized document size by number of topics of the document-topic distributions. This determines the percentage of words within a document that frequent topic k. As ibid. state, “Both LDA and STM can be framed as a factorization of the row-normalized D-by-V empirical word count matrix W, such that W  .”

The test the resulting topic quality of the models for exclusivity and coherence measures. Semantic coherence assesses how much the high probability words in a topic tend to co-occur together in documents, measuring how interpretable a topic is. Exclusivity measures how unique the words in a topic are compared to other topics. In other words, it indicates whether a topic contains words that aren't frequently found in other topics. The exclusivity scores in my model are generally less negative, suggesting that most topics in the model are relatively distinct, with unique sets of words. The semantic coherence scores are all positive and relatively high, indicating that most topics are coherent. The topics are likely to be meaningful and interpretable, with words within each topic co-occurring frequently.

The table below summarizes the most frequent words across topics in the STM using different methods for selecting and interpreting the most representative words in a topic. FREX (Frequency and Exclusivity). FREX is useful for identifying words that are uniquely important to a topic, helping to distinguish it from other topics.

I estimate the effect of the scale of withdrawing firms as a covariate on the distribution of topics in my STM analysis. The results reinforce previous findings from the LDA models in the main paper; where topic 3 (which centres around support for refugees, including services, government aid, and digital learning, indicating a focus on humanitarian aid and education is more likely to be associated with 'digging in firms', while 'withdrawing' firms are less likely to use these terms. This suggests firms that chose to dig-in tried to shore up against potential reputational losses from remaining in Russia by deflecting attention on their humanitarian efforts in Ukraine. Keeping in mind that in the total sample of firms, firms that remained where less likley than withdrawing firms to donate or offer humanitarian aid. The statistical significance of remaining firms being more

| Topic | Highest Prob | FREX | Lift | Score |
|---|---|---|---|---|
| 1 | busi, continu, compani, product, sanction, will, activ | materi, revenu, raw, necessari, total, reduc, less | buffer, finish, marin, selfsuffici, wood, common, decreas | lower, raw, revenu, finish, bureau, materi, buffer |
| 2 | support, donat, employe, continu, peopl, humanitarian, will | patient, medicin, food, relief, red, health, cross | antibiot, asthma, cardiovascular, crown, diseas, dose, generous | patient, froze, medicin, vaccin, clinic, trial, johnson |
| 3 | support, refuge, help, provid, donat, servic, govern | intesa, cloud, digit, learn, ukrain, read, educ | doctor, grassroot, italian, prize, recipi, ride, ridehail | arrang, intesa, cloud, skill, prize, web, school |
| 4 | support, will, peopl, work, continu, colleagu, also | client, member, firm, media, content, world, profession | baker, cruis, cyberattack, destruct, enact, forum, helplin | vendor, client, featur, content, stori, nonprofit, mandat |
| 5 | oper, employe, support, busi, will, invest, peopl | joint, ventur, hotel, equinor, exit, invest, oper | forese, imposs, deplor, earli, forefront, heat, investor | earli, equinor, rosneft, restaur, hotel, ventur, franchise |
| 6 | busi, compani, oper, employe, group, support, colleagu | crude, gas, store, oil, contract, renew, franchis | coat, crude, denmark, gap, north, pair, pipelin | crude, fritz, oil, gas, franchis, disinfect, document |
| 7 | right, human, busi, risk, respect, employe, oper | human, right, dilig, risk, assess, conduct, standard | actual, adequ, systemat, bulgaria, conflictsensit, depart, dual | human, viabl, dilig, esg, right, sap, screen |

**Table 22.** Top Words in Each Topic

likely to use these terms despite being less likely to issue donations/humanitarian relief, lends further credence to the argument that these firms utilized particular signalling strategies to deflect attention. Topic 2, which again revolves around support, donations, and humanitarian efforts, is much less likely to appear in withdrawing or suspending firm statements than firms that remained.

Topic 4 includes references to professional services, media, and people's work. It encompasses discussions related to supporting colleagues and professionals, along with work-related activities. Interestingly, withdrawing firms are more likely to use these terms, while digging in firms have negative effects. This might suggest that withdrawing firms are showing concern over assisting remaining staff members during their exit.There are smaller effects for companies across the topics, and date of withdrawal has no meaningful effect on differences in statements.

| Variable | Estimate | (SE) |
|---|---|---|
| (Intercept) | 0.130*** | (0.038) |
| Grade First (Digging In) | 0.087 | (0.061) |
| Grade First (Scaling Back) | 0.062 | (0.071) |
| Grade First (Suspension) | -0.005 | (0.050) |
| Grade First (Withdrawal) | 0.027 | (0.054) |
| First Action Numeric | 0.0006 | (0.001) |

**Table 23.** Topic 1 Regression Results

| Variable | Estimate | (SE) |
|---|---|---|
| (Intercept) | 0.322*** | (0.042) |
| Grade First (Digging In) | -0.111 | (0.063) |
| Grade First (Scaling Back) | -0.134 | (0.070) |
| Grade First (Suspension) | -0.182*** | (0.052) |
| Grade First (Withdrawal) | -0.229*** | (0.055) |
| First Action Numeric | -0.0006 | (0.001) |

**Table 24.** Topic 2 Regression Results

| Variable | Estimate | (SE) |
|---|---|---|
| (Intercept) | 0.069** | (0.026) |
| Grade First (Digging In) | 0.067 | (0.046) |
| Grade First (Scaling Back) | 0.009 | (0.049) |
| Grade First (Suspension) | 0.004 | (0.034) |
| Grade First (Withdrawal) | -0.051 | (0.034) |
| First Action Numeric | -0.0001 | (0.001) |

**Table 25.** Topic 3 Regression Results

| Variable | Estimate | (SE) |
|---|---|---|
| (Intercept) | 0.153*** | (0.041) |
| Grade First (Digging In) | -0.0608 | (0.061) |
| Grade First (Scaling Back) | -0.0336 | (0.072) |
| Grade First (Suspension) | 0.101* | (0.051) |
| Grade First (Withdrawal) | 0.272*** | (0.058) |
| First Action Numeric | -0.0012 | (0.001) |

**Table 26.** Topic 4 Regression Results

| Variable | Estimate | (SE) |
|---|---|---|
| (Intercept) | 0.165*** | (0.039) |
| Grade First (Digging In) | -0.0881 | (0.056) |
| Grade First (Scaling Back) | -0.0886 | (0.068) |
| Grade First (Suspension) | -0.0375 | (0.049) |
| Grade First (Withdrawal) | 0.0018 | (0.056) |
| First Action Numeric | 0.0016 | (0.001) |

**Table 27.** Topic 5 Regression Results

| Variable | Estimate | (SE) |
|---|---|---|
| (Intercept) | 0.0471 | (0.030) |
| Grade First (Digging In) | 0.0204 | (0.049) |
| Grade First (Scaling Back) | 0.0476 | (0.058) |
| Grade First (Suspension) | 0.0915* | (0.039) |
| Grade First (Withdrawal) | 0.0613 | (0.043) |
| First Action Numeric | -0.0004 | (0.001) |

**Table 28.** Topic 6 Regression Results

| Variable | Estimate | (SE) |
|---|---|---|
| (Intercept) | 0.1127** | (0.037) |
| Grade First (Digging In) | 0.0849 | (0.058) |
| Grade First (Scaling Back) | 0.1387* | (0.071) |
| Grade First (Suspension) | 0.0284 | (0.049) |
| Grade First (Withdrawal) | -0.0826 | (0.049) |
| First Action Numeric | 0.0001 | (0.001) |

**Table 29.** Topic 7 Regression Results

# Who Needs To Be Seen To Be Green?

# How Reputational Pressure Affects Responses To Climate Change

Hayley Pring

Paper Three

**Abstract**

Domestic and international laws have been slow and ineffective at regulating firms in response to climate change. For this reason, voluntary self-regulation and private regulatory bodies have become increasingly important. Despite claims that firms oppose costly carbon mitigation strategies, previous research has shown variation in the extent to which firms signal their performance on climate change. What explains this variation? I develop a four-pronged theory of reputational pressure to answer this question. First, firms face different reputational pressure based on their primary consumer audience. Business-to-consumer firms face greater reputational costs of failing to act on climate, and therefore avoid exposure. Second, 'Climate-conscious' reputations matter when consumers are primed to value them. I argue that climate disasters induce consumer awareness, and firms respond by signaling to audiences their climate-reputation. Third, reputations have cascade effects. Firms are more likely to move to disclosure following large early-movers within their sector. Fourth, there is a trade-off for firms signaling action on

climate; attention brings greater scrutiny, which may damage the firms reputation more than keeping quiet. Following poor climate evaluations, firms become much less likely to disclose in the future. Using Australia as a case study, I employ data on firm disclosure and climate-mitigating action from the Carbon Disclosure Project between 2010-2021. In addition to natural disaster data from EMDAT, I proxy for the material effects of natural disasters using insurance claims, and use a difference-in-difference estimator to compare high and low years following climate disasters. The empirical tests show reputational pressure can both induce and paralyze firm action on climate.

# 1 Introduction

A central message that emerged from the 2015 Paris Climate Conference, and subsequent COP meetings, is that the implementation of intergovernmental climate solutions will require not only cooperation from governments and cities, but also private sector actors taking on "global governor" roles (Avant, Finnemore, and Sell 2010). This call to action represents the political and litigious barriers of regulating industries domestically across different sectors. For example, some sectors involve political constituencies that are highly sensitive to the visible costs of policy interventions on ubiquitous carbon-emitting products, such as gasoline. In other sectors, industrial production revolves around acutely competitive commodities such as steel, and here firms are capable of organizing to hamper, delay or kill policies that could harm price-sensitive and trade-exposed industries (Cullenward and Victor 2020). As a result, designing and implementing regulations for change is complex and difficult. Without regulatory requirements, firms have many reasons not to act on decarbonization. Least of which are the acute and immediate costs of taking action, as well as the logistical issue that key technologies necessary for decarbonization may be immature or risky to adopt (Victor, Geels, and Sharpe 2019). Yet, despite the fact that decarbonization and climate mitigation processes impose substantial costs on corporations, in 2018, 2,138 firms from 145 countries, amounting to roughly $36.6. trillion USD in revenue, voluntarily pledged to act on climate change (Hsueh 2017). As part of this, a growing trend has witnessed the cooperation of firms with private-led quasi-regulatory bodies towards implementing internal change. This has prompted a puzzling observation; why do some firms choose to cooperate with quasi-regulatory bodies on climate change?

I argue that reputational concerns can drive firms to both avoid and signal

support for internal climate regulation. First, firms face heterogeneous reputational pressure to present as climate-conscious; firms with large consumer audiences (consumer facing businesses, or B2C) face greater scrutiny than business-to-business (B2B) facing firms. Rather than this motivating action, these firms avoid alerting audiences and conceal themselves from scrutiny. In line with existing research, the social pressure to project a climate-conscious reputation remains low as consumer demand is hampered by the cognitive dissonance created by hyperbolic discounting. However, incentives to build or maintain a climate-conscious reputation increase when audiences are 'primed' to scrutinize firm behavior. Natural disasters linked to climate change can create these priming conditions, triggering firms to project climate-conscious reputations in the short-term. Consumer facing firms, and firms linked to higher emissions, face greater reputational pressure to act. Third, reputations have negative externalities. Firms associated with sectors that are linked to 'high emissions' are more likely to act on internal climate regulation to safeguard against a negative collective reputation. Fourth, reputations have cascade effects; large 'early-mover' firms can induce a cascade of similar behavior at the sector-level.

The literature on why it's difficult to regulate carbon emitters, and why some agents – in this case, firms - are incentivized to act has largely been split. In regards to the former, the account of climate change either emphasizes a collective action problem of overcoming free-riding, or a distributive problem whereby decision making in one space or time period affects actors at different points of space or time (Aklin and Mildenberger 2020). Collective action approaches (Barrett and R. Stavins 2003; Keohane and Victor 2016; Ostrom 1990; R. N. Stavins 2011; Nordhaus 2015; Budolfson et al. 2012; Victor 2011) and distributive politics offer two important points 1) climate change is a pub-

lic good, 2) but the harms are unevenly distributed (Gazmararian and Milner 2022). Both of these models of climate politics speak to the variation in actors support over space and time on the demand-side.

On the supply-side, work focused on the puzzle of why firms adopt or support climate mitigation has been dominated by two approaches. The first suggests support for regulatory measures aimed at climate change is a type of 'strategic accommodation' (Hacker and Pierson 2002). That is, firms support moderate regulation to avoid more costly measures in the future. Strategic accommodation features prominently in the literature on the disappearance of Big Oil's opposition to climate regulation in the 1990s (Kolk and D. Levy 2001; D. L. Levy and Egan 1998; Newell and Paterson 1998).[1] Running counter to this position is the view that the firm has economic interests in the regulation itself. There are perceived economic advantages for large incumbents to adapt to new technology and squeeze out smaller and emerging competitors. For instance, Kennard (2020) shows that firms support domestic climate change regulation to gain advantage over market competitors, where some firms possess an advantage in adjustment costs. Similarly Colgan, Green, and T. N. Hale (2021) introduce an asset revaluation theory whereby actors possessing a larger share of climate-vulnerable to climate-forcing assets should be more supportive of regulatory measures(ibid.). Both sides have offered keen insights into firm behavior, however supply-side arguments leave out stakeholder concerns and the social position of the firm in relation to their customers. Exclusively demand-side arguments overlook the variation in responsiveness from actors, or how sensitivity to these demands may vary.

[1] Also further support from US oil and gas majors following the Kyoto ratification, 2007 Supreme Court ruling granting the EPA authority to regulate green house gas emissions under the Clean Air Act, and new state laws (D. L. Levy and Kolk 2002; Meckling 2011; Pinkse and Kolk 2012; Pulver 2007; Skodvin and Skjaerseth 2003; Skjærseth 2013).

To explain variation in firm engagement with voluntary regulatory bodies on climate change, I introduce a theory of reputational pressure that incorporates both demand and supply side theories for regulating climate change. Empirically, I leverage data from a rapidly growing private semi-regulatory body - the Carbon Disclosure Project (CDP). The CDP, a registered member of the UN SDG's Partnership Platform, operates a global environmental disclosure system. The program invites firms to submit reports and surveys which are then scored by the CDP to assess the extent to which the firm has a) disclosed its carbon emissions, and b) made progress towards carbon-mitigation. Firms may refuse to participate and, as a voluntary initiative, the CDP has no mechanism through which they can enforce compliance from firms. Since non-compliance has no material cost, the CDP is a suitable case study to evaluate the role of reputation as a mechanism in shaping firm behavior.

Focusing on Australia as a case study, I collect disclosure and scoring data between 2010-2021. Australia is an ideal setting for two reasons. First, it is exposed to visible and negative shocks from climate change and due to the demographic clustering of the population, natural disasters have widespread, direct effects on the public. Second, firms have been required to submit reports on carbon emissions to the Clean Energy Regulator since the 2007 National Greenhouse and Energy Reporting Act (NGER Act). Mandatory reporting requirements by all publically-listed companies reduces the selection bias of high-performing firms.

I combine annual firm-level data from Refinitiv Eikon, with data on climate-linked natural disasters between 2009-2020 from the EMDAT dataset. I find

that consumer-facing firms are less likely, on average, to disclose than B2B firms. Second, following natural disasters, firms are much less likely to disclose information, however firms that do disclose score much higher following hig-disaster years than low-disaster years. Third, following a large 'early-mover' in an industry, smaller firms are more likely to cooperate with the CDP. Fourth, high intensity emitting sectors that are consumer facing score much higher than low-intensity firms. Lastly, poor performing firms that received failing grades are much less likely to disclose information in the future. I find substantial support for the role of fluctuating reputational pressure in inducing firm disclosure and implementation of carbon-mitigation strategies.

In addition to using a variety of model specifications, and estimation measurements, I address the potential issues of variation in the scale of natural disasters. To capture the real impact on consumer audiences of natural disasters that triggers climate awareness, I use a proxy for the potential total economic losses from natural disasters linked to climate disasters by using household insured losses from the Insurance Council Australia database. To address confounders and endogeneity issues, I use a stacked difference-in-difference estimator to compare disclosure and scoring following high versus low disaster years.

This research makes three contributions to the study of businesses in environmental politics. First, it sets up a probabilistic explanatory model of firm behavior on climate change. Existing explanation focus on comparative advantage of low versus high carbon emitters, or theories of hedging out the competition, rooted in economic interest and focused on within-business dynamics. In contrast, a reputational theory of firm behavior builds on these ideas and incorporates a view of heterogeneous audience types (consumers versus businesses),

with different expectations over firm behavior (low or high emitters), depending on fluctuating awareness and preferences over firms adopting climate-conscious reputations (hyperbolic discounting and climate 'shocks'). The through-line in the theoretical explanation is that firms face varied and shifting pressure to project climate-conscious reputations, but also that supply-side dynamics can help drive reputational dynamics, rather than arising exclusively from audiences. Second, I incorporate ideas from behavioral science largely used in the context of individuals to understand firm behavior as having a 'group dynamic' where concern for reputation occurs not in a vacuum, but in response to competitor actions. Third, I develop original findings on which firms and industries are more inclined to signal action on climate, and which abstain, and how negative evaluations can deter future firm participation in voluntary semi-regulatory schemes.

These findings build on more current work showing that firms and industries that intensively emit carbon are more likely to actively oppose climate action depending on their position in the supply chain (Meckling 2015). While this is the case, intensive emitting firms and industries face greater reputational pressure to project climate-conscious reputations. This suggests a gap between the actions of the firm, and the image they project. This largely confirms the position of many environmentalist who are skeptical of voluntary regulations (Steinzor 1998), on the basis that these programs are used by firms to "greenwash" an otherwise poor environmental performance (Prakash and Potoski 2006; King and Lenox 2000). Though possible that these programs are used to project particular environmentally-conscious reputations, voluntary programs can influence firms into implementing progressive environmental action, resulting in pollution reduction and higher compliance rates with domestic regulations (An-

ton, Deltas, and Khanna 2004; Dasgupta, Hettige, and Wheeler 2000; Potoski and Prakash 2005b; Potoski and Prakash 2005a).

## 2 The Carbon Disclosure Project: Background

Climate legislation has been slow, patchy, and inadequate in most high-emitting nations (Keohane 2015). Where it has been successful it has offered small rewards after years of negotiations (Eskander and Fankhauser 2020). In the absence of satisfactory legislation, private environmental regulatory bodies can fulfil the role of governing institutions by providing a coordinating mechanism to monitor, set standards, provide enforcement, and create ad-hoc dispute settlement mechanisms (Vandenbergh, Rossi, and Faucher 2020). Among the most impactful of these private environmental governance initiates are voluntary semi-regulatory programs that allow firms to disclose information, and receive ratings on their performance. One notable example is the Carbon Disclosure Project (CDP). As a global disclosure system, the CDP invites firms around the world to be surveyed and submit reports on carbon emissions and internal firm regulations towards climate change on an annual basis. Founded in 2000, the CDP is a major private regulatory body with 13,000 companies representing 64% of global market capitalization disclosing to the CDP in 2021.

Submissions are voluntary and firms are offered three different disclosure choices. The firm can 1) choose not to participate or respond, 2) answer with confidentiality, where answers are hidden from the general public, and 3) answer with public disclosure. The CDP then surveys each firm, assigning a grade that scores the firm on its efforts towards 'environmental stewardship' across three programs: Climate Change, Water, Forests and Supply Chain programs. Three characteristics of the CDP make it an ideal case study for evaluating

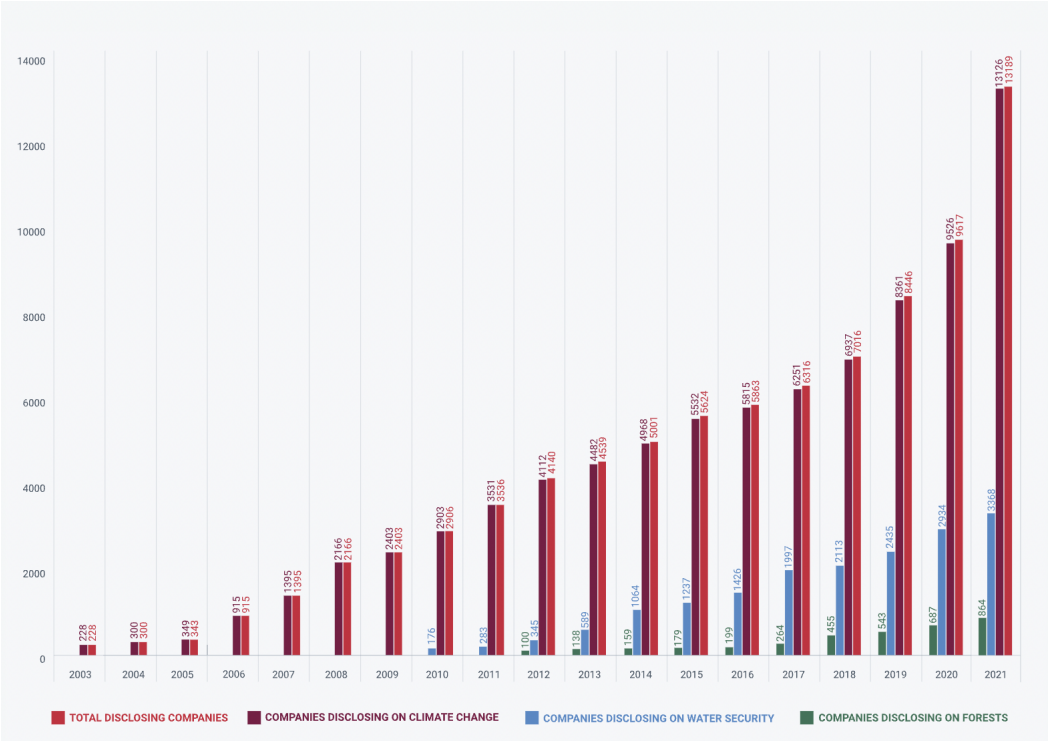



Figure 1: Disclosures issued by Firms to the CDP (2003-2021). Data source: https://www.cdp.net/en/companies/companies-scores.

the role of reputation on firm behavior. First, as a voluntary initiative, the CDP resembles the structure of other private regulatory bodies. The scoring is based solely on the activities and positions disclosed to the CDP. The CDP does not set carbon targets, and a major criticism levied by skeptics is that it does it have the capacity to verify the information disclosed by firms (Matisoff, Noonan, and O'Brien 2013). Second, there are no material benefits to joining the CDP, other than the advertisement of the firm's assessed reporting on its climate change activity should it receive a high grade. In this way, the CDP can be utilised by firms to project particular reputations, since "the CDP gives firms a way to distinguish themselves from competitors and gain recognition for going beyond compliance" (ibid., p. 289). Third, the CDP scores firms on two

metrics; whether or not they respond to requests to disclose information, and the information itself. Thus, firms that do disclose know they may be subject to a negative evaluation, creating a trade-off for firms in choosing to participate. Positive evaluations can help the firm's reputation, while negative evaluations may harm them.

## 3 Variation in Firm behavior on Climate Change

In developing an argument on the role of audience reputational pressure in shaping domestic firm preferences over how they voluntarily engage with regulatory bodies over internal carbon regulation, I straddle two literatures on firm preferences in political economy and the behavioral science of climate policy.

Whether or not firms or industries choose to act, and signal action, on internal carbon regulation and carbon-mitigation strategies, will be influenced by a confluence of social pressure, political and legislative change, and the distributive effects of climate policy (D. L. Levy and Kolk 2002; Meckling 2015). The first category of determinants in the political economy literature focuses on how economic factors shape firm policy preferences. Recent research has highlighted the differential effect of trade exposure on polluting sectors (Genovese 2019), and the competitive gains for large domestic firms against foreign or emerging competitors (Kennard 2020). Adjustment costs for firms also varies among sectors where the availability of mature technology for low-carbon innovation differs across industries, and non-harmonization across jurisdictions exposes businesses to costly moves in external markets (Bumpus 2015; Cullenward and Victor 2020).

A second category of factors have emphasized the conditioning effect of po-

litical and institutional contexts that different firms and industries are situated in. Firms located in industries that are large emitters of greenhouse gases are likely to oppose costly decarbonization measures as they face steep mitigation costs (Goel 2004; Markussen and Svendsen 2005; Cho, Patten, and Roberts 2006; Martin and Rice 2010; Kim, Urpelainen, and Yang 2016), and the same holds for firms that consume electricity linked to fossil fuel sources (Cheon and Urpelainen 2013; Kelsey 2018; Kennard 2020). Firms located within a supply-chain may face downstream costs of carbon-mitigation efforts. The distributional consequences of regulatory policies are greater for large-emitters, who are likely to invest greater resources into lobbying to oppose such changes (Colgan, Green, and T. N. Hale 2021), or adopting 'strategic accommodation' stances on moderate regulatory policies to prevent more drastic changes (T. Hale 2011). In contrast, more recent research has found that some large firms are not feigning support for regulatory change. Major oil firms have invested in natural gas with the expectation that an eventual moderate carbon price will increase the competitive ability of natural gas to become the predominant base-load partner of intermittent renewables and a transitionary fuel towards decarbonization (Skjærseth 2013).

This article integrates these insights, but focuses on this third set of social factors that influence firm behavior; the role of social forces and internal firm characteristics in influencing firm action on internal climate mitigation strategies. This work points to external influences like strategic competition between interest groups, the strength of supporters and opponents' coalitions (Cheon and Urpelainen 2013), and stakeholder pressure (Martin and Rice 2010). The focus of the literature on external versus internal dynamics has left out an important part of the story on firm responsiveness to climate change – the interaction be-

tween consumer and stakeholder preferences and firm behavior, and the extent to which the firm is exposed to sharp shifts in consumer attention to a firm's environmental reputation.

It is difficult to speak about how social influences can pressure firms to implement internal climate regulations without considering what motivates the consumer to care about the climate. It is this second set of agents - the audience and customers of the firm - that necessitates bringing in discussions on the nature of consumer preferences over climate change. A difficult premise for climate change is that people often make decisions that leave them worse off. Human behavior is both spatially and temporally context-dependent. When the negative effects of climate change are unevenly distributed across time and space, individuals engage in hyperbolic discounting - meaning "the time-declining discount rate to trade-offs between present and future consumption" (Hepburn, Duncan, and Papachristodoulou 2010, p. 190).[2] Although this applies to costly changes that firms may make, or increased downstream costs for consumers unwilling to pay more for 'green products', it also applies to attention. While scientific evidence and discourse over the increasing threat of climate change has increased over the last decade, constant exposure to the same stimulus leads to habitation (Weber and E. J. Johnson 2009), resulting in desensitization to risk evaluation. Taken in conjunction with the dispersion of responsibility of emissions and climate governance (individuals, corporations, governments and international actors), consumer attention is unlikely to be orientated and collectively organized enough to impact whether firms need to make costly investments in, and signal, climate-conscious reputations. Thus, when the environment remains in a constant state, broadly speaking, the effect of social pressure from consumers on firm behavior may be low. Changes in the environment that trigger sharp

[2]See also: Weber and Stern (2011).

'shocks' can alter this equilibrium, meaning that consumer pressure for firms to signal climate-conscious reputations varies in response to fluctuating awareness. The next section introduces a theoretical framework for the effect of reputation on firm responsiveness to climate change (Posner, Rothbart, et al. 2007).

# 4 A Reputational Theory of Firm-Level Climate Responsiveness

Reputations are "a perceptual representation of a company's past actions and future prospects that describes the firm's overall appeal to its key constituents when compared to other leading rivals" (Fombrun 2005, p. 72). The main reputational currency of firms is quality, with research showing positive reputations are correlated with better return on assets (Benjamin and Podolny 1999; Standifird 2001; Boyd, Bergh, and Ketchen Jr 2010; Deephouse 2000). Aside from economic performance, good reputations leads to better employee candidates and customers (Turban and Cable 2003), and the ability of the firm to deviate from strategic behavior without penalty (Deephouse and Carter 2005). Certain firm actions can generate reputational losses, such as downsizing (Love and Kraatz 2009), or reputational gains, such as when firms improve product quality. Increasingly, firms use Corporate Social Responsibility (CSR) Reports to build their reputations (Esen 2013; Sethi, Martell, and Demir 2016). The necessity of cultivating certain reputations, such as quality of products and services, are fixed, while the impetus to develop reputational types (such as human-rights abiding) grow in reaction to changes in consumer preferences.

Firms face pressures to signal reputational types and distinguish themselves from competitors; if consumers develop preferences towards greener corporate

policies, firms will respond by signaling their pro-environmental policies. There is some evidence that this occurs at the individual-level; where the desire to develop an altruistic reputation motivates individuals to purchase green products in public (though not privately) despite the higher price and lower quality of the goods (Griskevicius, Tybur, and Van den Bergh 2010). While the visibility of the firm effects the extent to which the firm will feel pressure to send such a signal, the insight can be transposed to private-sector actors.[3] Of course, firms may achieve this by releasing environmentally-friendly products - for instance McDonalds substituting paper wraps for styrofoam containers (Stafford and Hartman 1996), however firms can send clearer signals by cooperating with semi-private regulatory bodies who independently evaluate firm performance. The necessity to signal a climate-conscious reputational type is dependent on 1) the institutional structure of the firm, and their sensitivity to consumer audiences, and 2) the extent to which the consumer audience is primed to care about this reputation.

An external factor that can trigger pressure on corporations to respond to, or signal action on, climate change is the proximity of the firm to the consumer. Consumer-facing firms are companies which "supply goods or services directly to consumer markets instead of supplying to another business entity" (J. Haddock-Fraser and Fraser 2008). In contrast, firms which predominantly involved in business-to-business (B2B) activity have a lower level of interaction with public consumers of their products. As Kolk and Pinkse (2007)(374) note, "the position of the company in the supply chain determines to which extent a company depends on its customers and thus follows a concomitant climate strat-

[3]Hollyer, Rosendorff, and Vreeland (2015) make a similar point on how information disclosure by autocratic governments affects public mobilization and protest, leading to greater instability. In their model, invoking reputation is not necessary because the actor disclosing information has no peer comparison group.

egy". Several studies find a positive association between the proximity of the firm to the consumer and pro-environmental activities (Damert and Baumgartner 2018; J. E. Haddock-Fraser and Tourelle 2010), as well as carbon disclosure (J. Haddock-Fraser and Fraser 2008). According to these accounts, reputational pressure increases with the firms position in the supply-chain vis-à-vis the final consumer. Although these studies have either focused narrowly on one industry, or have selected case studies of top-performing firms, and it is unclear how impactful this factor is.

While the institutional design of the firm should affect the sensitivity of the firm to reputational costs from consumer audiences,[4] I argue that the effect of this reputational pressure is mediated by 1) the extent to which the firm is cognitively associated with high GHG emissions from consumer audiences, and 2) the trade-offs of signaling firm action on climate change. First, consumer facing firms that are not linked to high-emissions will face lower incentives to project a climate-conscious reputation. Second, B2C facing firms face higher levels of public scrutiny, and submitting to regulatory bodies can increase attention on the firm by increasing the reputational costs of submitting and receiving a poor performance evaluation. Thus, I expect reputational pressure to produce two different outcomes for consumer facing firms. Higher levels of scrutiny can dissuade firms from submitting to regulatory bodies. However, consumer facing firms that are high-emitters already face negative public scrutiny, and can use CDP scoring to repair their reputation; these firms are more likely to submit than high-emitting B2B firms. B2B firms face lower levels of reputational pressure and therefore will be more likely to submit, but also score lower.

[4] Firm's face different incentives to signal different information based on who their audience is. Just like the role of reputation in firm's developing reputation for quality; consumers use this information to distinguish between types of firms and make decisions on who to engage with. The reputation dimension is important for observer's to make comparisons and then make decisions on who to engage with, contemporaneously and in the future.

One example of the transparency issue with B2B facing firms and their relationship with climate disclosures is Aurizon Holdings. Aurizon is Australia's largest rail freight operator, and an ASX publicly listed company. Aurizon is an interesting case since it indirectly contributes to Australia's largest emissions base; the largest source of Australian emissions is its power and heat sector, which account for over 40% of GHG emissions, and more than half of electricity is generated by coal. Aurizon generates around 80% of its revenue from hauling coal around the country. In terms of direct emissions, in 2020-2021 it was the 73rd largest GHG emitting controlling corporation in Australia. The responses submitted by this company in reaction to requests for disclosure by the CDP are interesting. The company submits to the CDP for review between 2013-2019, but following Australia's black summer, the company stops submitting. The assumption here might be that the company scored well in CDP reports from 2013, and were encouraged to keep submitting. This is not the case, in fact in 2013 and 2014 the company had failing grades (E, D). Yet, if you look at the annual reports submitted by the company, Aurizon dedicates an entire paragraph to the company's participation in CDP, and then remarks, "In November 2013, Aurizon was recognized as a finalist in the CDP ASX 200 Climate Leadership Awards for 2013. The finalist nomination was in the category for 'Best New Responding Company' which recognizes first-time ASX 200 companies with the highest CDP climate disclosure scores for that year". The company does not mention that it scored a failing grade. As a B2B firm, the company's performance is relatively hidden, for shareholders or other company's, their reference point for the company's performance on environmental issues is their annual report. As a B2B Aurizon continues submitting despite its poor grades, but this practice ends following the black summer, when attention on climate-behavior

increases.

For a contrasting example of B2C companies, and their fear of poor grades, we can look at the example of the Arden Leisure Group. The company submits between 2012-2016, but after receiving three years of failing grades, in 2017, the company submits but requests for non-disclosure of its scoring. This is a process by which companies can be evaluated without the evaluation being public; in some ways it can allow the company to 'test the waters' so to speak, of where the company stands without having those results publicized. After 2017, the company stops responding for requests to submit. Another case is GPT Group, a real estate and investment trust in Australia. The company submitted 2012-2018, after declining from a C in 2017 to a D failing grade in 2018, the company stops submitting and opts to 'decline to participate' rather than simply ignore the request. Another company, GrainCorp, fails to participate from 2011-2012, in 2013 they participate but under conditions of non-disclosure, after this the group continues to not respond for request for disclosure. Since the 2013 disclosure report is not scored and is confidentially held, it's only guesswork to say the result was enough to discourage the company from submitting in the future, lest their poor scores be revealed.

I therefore introduce both the moderating effect of carbon emitting intensity and disclosure versus scoring into my initial predictions:

> Hypothesis 1a: Consumer facing firms will be less likely to disclose for evaluation to the CDP, but B2C firms that do disclose will score higher relative to B2B firms.
>
> Hypothesis 1b: B2C firms that are also high-emitters will be more likely to score higher on CDP reports.

## 4.1 Natural Disasters

As previously stated, individual behavior is context-dependent. Awareness of risks is influenced greatly by the availability heuristic (Renn and Rohrmann 2000); if a certain risk is cognitively "available" - present, felt or salient - then individuals will have a heightened response to that risk (Sunstein 2006). Here I argue that consumer audiences don't have fixed preferences over climate change, they fluctuate when climate events 'prime' them and reverse their preferences, increasing awareness and attention to whose to blame for climate-linked natural disasters.

A bi-product of a climate-linked disaster is to trigger greater awareness of the relative contribution of firms to the adverse state of the climate. A reputation for being environmentally-conscious is less important in equilibrium where the effects of climate are difficult to tangibly link back to private sector actions, but when the climate conditions are adverse, bad climate policies may garner more attention. Firms are better adept at 'blending in' where bad climate policies are difficult to distinguish between individual firms within an industry.[5]

Studies have shown a sharp increase in awareness of climate change following climate-linked disasters with tangible effects on behavior (Jetten et al. 2021; McAllister and Oslan 2021). Baccini and Leemann (2021) find the exogenous shock of a climate-linked disaster (flooding) primes voters in Switzerland in favor of climate protection. Climate is generally second order in political debates, yet other studies find significant effects on voting behavior in general elections where climate is rarely salient (Bechtel and Hainmueller 2011; Gasper

[5] See Grenadier, Malenko, and Strebulaev (2014) on the increasing importance of reputation during adverse 'shocks': agents delay the abandonment of R&D projects or investments during periods of economic shocks when it's less likely to send a negative signal of individual ability; thus 'blending in' with the crowd.

and Reeves 2011). Hazlett and Mildenberger (2020) show that natural disasters can help overcome the temporal mismatch between short-term costs and long-term benefits of climate policy change. The authors find that exposure to wildfires increased support for climate-related ballot measures for individuals living within the vicinity of the event.

Although climate change, and environmentalism in general, are salient political and social issues, agents generally suffer from hyperbolic discounting and cannot make current trade-offs for future benefits. In situations where agents experience direct and consistent effects of climate change, effects of decisions are more tangible and consumers may be willing to pay a premium for firms that possess an environmental reputational type. Studies suggest reputation can be leveraged for competitive advantage, finding consumers use aggregate information to form reputational judgments which subsequently influences their purchase decisions (Graham and Bansal 2007). Thus, as the negative effects of climate change increase, individuals may accept some loss of material utility by choosing firms with climate-conscious reputations to reduce their carbon footprint. The shift in consumer preferences and climate awareness cultivates a culture that values reputations for environmentalism; firms respond to this shift in preferences by increasing their attention to projecting climate-conscious reputations. However, the costs of increased attention on firms' environmental performance can dissuade them from disclosing information and increasing negative scrutiny. I therefore expect two polarising outcomes:

> Hypothesis 2a: Firms will be less likely to disclose following a year of high climate-linked natural disasters.

> Hypothesis 2b:Firms that do disclose will score higher on CDP evaluations in years that follow a high number of climate-linked natural disasters.

These theories examine the role of reputation across sectors, however consumers are more aware of the complicity of some sectors over others, and reputations are often 'bench-marked' to an agent's peers rather than firms in general. I introduce two novel concepts of 'cascade effects' and 'reputational externalities'.

## 4.2 Reputational Externalities

Firm reputation is not the sole product of a firm's actions. For all agents, observers typically use temporal context to make judgments or inferences about a firms current actions. The historical context of an agent's actions allows observers to assess relative, rather than absolute, improvements or declines. Similarly, there is also a comparative context among firms within the same industry, creating the effect of *reputational externalities*. The notion of reputational externalities explains how an agent's behavior can threaten the collective reputation of the group (Barnett 2007; Tirole 1996). We can conceive of reputational externalities to be a situation where "the actions of one group may damage the reputation of another group, thereby reducing overall welfare" (Evans and Guinnane 2006, p. 2). This phenomena is discernible in IR; for instance when states join together under an organization or grouping, individual behavior of one agent can damage the reputation of other members. Or, perhaps, when heterogeneous agents or bodies develop common reputations through associations or long-standing alliances. Political parties are particularly conscious of the effect of individual member's behavior on the collective reputation of the party (Marland and DeCillia 2020; Kreps 1984; Calvert 1987).

The idea is more prevalent in the business and organizational literature where reputational spillovers have been observed in firm layoffs (Goins and Gruca 2008), after corporate social responsibility (CSR) crises (Winn, MacDonald, and Zietsma 2008), and geographical proximity of firms (Bertels and Peloza 2008). The adoption of environmental policies in particular suggests significant reputational externalities are driving firm behavior. Hargrave (2008) finds that whether or not firms participate in 'green clubs' is influenced by peer-firms. A statement by an Amoco executive in 2001 following an Exxon oil spill illustrates the reputational pressures that intra-industry firms can exert on each other.

> *"We are still an oil company, and we still have to live with the sins of our brothers. We were doing fine until Exxon spilled all that oil. Then we were painted with the same brush as them."*(Hoffman 2001)

Observer perceptions of an industry sector are conditional on the reputations of firms within that sector, as evidenced by findings in studies by Shamsie (2003), Cable and Graham (2000), and Brammer and Pavelin (2006). Intra-industry firm reputation is damaged when other firms within the sector, who provide similar services, possess poor climate or environmental policies. Firms will push for stringent regulatory regimes to create a common reputational climate practice among heterogeneous firms. In this way, concerns over the collective reputation of the industry drives firm support for improving collective reputation through regulatory regimes (Standifird 2001).[6] As a high-emitting sector with a notorious link to carbon emissions, I focus on the oil, gas and mining industry to determine whether within-sector negative externalize motivate

[6]One might argue that some industry sectors, such as oil, will never be able to develop 'good' environmental reputations. However, it's more important to firms to avoid negative reputations than it is for them to cultivate 'good' reputations, where "negative reputations are more impactful and deterministic than positive reputations"

firms to disclose and self-regulate:

> Hypothesis 3: Firms in the oil, gas and mining sector will be more likely to disclose information to the CDP.

## 4.3 Reputational Cascades

Economic self-interest might predict that firms will support climate policy when such a policy benefits the industry. However, self-interest cannot explain why, in situations where the regulatory proposal imposes net costs, firms that once held opposing positions suddenly reverse their stance (Vormedal, Gulbrandsen, and Skjærseth 2020). And, although the conventional 'strategic accommodation' view can offer insight into why firms might offer support for climate-mitigation proposals as a means to offset radical alternatives, the theory is less adept at explaining the heterogeneity in firm positioning or the timing in which firms switch positions. One mechanism through which reputation can incentivize firms to cooperate with voluntary private-regulatory bodies is a *cascade effect*.

Reputations are not formed in vacuums, and they rely on a tenuous informational environment. Reputations are qualities that actors do not own – rather "because credibility and reputation are relational concepts they depend entirely on the perception of others" (Jackson 2016, p. 16). Observers form reputational judgments by assessing an actor's alike peers to benchmark them to or against; and actors assess their reputational type based on their peers as well. Thus, whether or not an actor has a 'good' or 'bad' reputation (their reputational type), depends on the actor's peers. Meaning that an actor may make an action or abstain from an action based on what their peers do (or do not do). This is called a cascade effect.

When making judgments about facts, values, tastes or even decisions, cascade effects can occur. They can happen privately or publicly – within or between individuals, groups, or institutions. Effective connections between alike actors increases the likelihood of a cascade effect. Building on the definition offered by Sunstein (2019) work on social cascades, I define reputational cascades as "large-scale judgments and movements in which actors begin to believe something, or do something, because of the beliefs or actions of a few early movers". First-mover advantage is a concept recognized in non-cooperative game theory, but it also has an interesting effect on group-dynamics by polarizing outcomes or leading to inefficiencies based on a tendency to defer to initial judgments (Kahneman, Sibony, and Sunstein 2021).

In a reputational cascade, actors may move against their preferences in a conformity-rewarding process. Reputational cascades are a distinct phenomenon to informational cascades, but we can think of them as siblings. Here is the distinction: in informational cascades, individuals fail to reveal private information to the group out of concern that their information is 'wrong' if their information contradicts that of the first early-movers. In contrast, in reputational cascades the failure to reveal private information to the group does not arise out of a fear of that information being 'wrong' but rather the fear of exclusion or disapproval for contradicting the group. Kuran (1997) refers to this as 'pluralistic ignorance'. Most of the work on cascades is based on individuals and has been tested in laboratory settings (Anderson and Holt 1997; Hung and Plott 2001; Willinger and Ziegelmeyer 1998). The idea of these early-movers and cascades can clearly be seen in different settings.

The concept of 'early movers' is foreign to IPE and the literature on climate change, despite how important it is for shaping policy. Early-movers are often acting on private information and therefore their decision can send a 'private signal' about whether or not they've made the correct judgment. However, when you have two early-movers sending the same signal, others may rationally fall into line, even if it's incorrect. As Anderson and Holt (1997, p. 847) note, "Initial misrepresentative signals start a chain of incorrect decisions that is not bro- ken by more representative signals received later".

As a result of this noisy signaling, though a reputational cascade is not a normative concept in and of itself, they can lead to "widespread errors, factual or otherwise" (Sunstein 2005). In some situations of IR, a cascade effect will overcome collective torpor or reluctance in creating public goods, for instance acting on climate change or environmental governance. We can see how this would play out in IR. Suppose a state suggests that the actions of state B in dumping toxic nuclear waste is a serious problem, and that state C concurs publicly with state A, not because state C necessarily agrees with state A but because they do not want to seem to state A as ignorant or indifferent to health or environmental protection. If we move down to the level of the firm, large early-movers can have a similar reputational cascade effect. Cascade effects among firms in relation to corporate social responsibility are rife. For instance, the reaction of multinational companies to the invasion of Ukraine by Russia in early 2022. Following the initial invasion, the corporate exodus out of Russia followed a pattern of within-sector closely linked firms following withdrawal strategies of their peers. Oil and energy companies largely moved within a week-long unison while discretionary goods – fast food restaurants and chains like Starbucks and PepsiCo – moved within days if not hours of McDon-

ald's announcement to leave. Firms that remained – or only partially withdrew or suspended operations like Papa Johns and Yum Brands! – were subject to within-group comparisons, since the early-movers had now shifted the relative benchmark through which consumers could evaluate their CSR reputations. In the area of climate change and environmental governance, cascade effects are evident across litigation, where environmental regulations at sub-global levels can trigger regulation at higher levels (Engel and Saleska 2005). A clear example of a reputational cascade causing firms to change their behavior in reducing emissions was clear when Xcel Energy of Minneapolis agreed to disclose the risk that global warming posed to its business in 2008, causing competing firms to follow suit. Within months Dynegy followed suit, and AES Corporation the following year.

These examples highlight how cascades can increase the reputational pressure for firms to conform. If we think about the preferences of firms as being represented by a utility function of a) the company's private utility and b) their reputational utility, the private utility is the firm's preferences over whether or not they disclose carbon, and whether or not they implement carbon-mitigating policies and internal regulations, while the reputational utility is how their decisions are viewed by consumers and shareholders. The reputational utility increases when competitors and within-industry peers publicly signal a move towards green-policy change. I therefore expect:

> Hypothesis 4: Firms will be more likely to disclose to the CDP if large-firms within their industry submit the year prior.

While cascades in H3 explain why firms may be motivated to move, H4 and H2 help explain the conditions under which firms restrain from disclosing.

# 5 Research Design

## 5.1 Case Study: Australia

Prior studies investigating the effect of carbon information disclosed by firms on variables such as shareholder wealth, pricing, and company market capitalization have been criticized for using voluntary disclosed carbon emissions data, given the issues with selection bias and endogeneity towards high-performing firms incentivized to submit this information voluntarily. However, Australia offers an ideal case study to alleviate these issues since Australian public firms have been required to submit reports on carbon emissions to the Clean Energy Regulator since the 2007 National Greenhouse and Energy Reporting Act (NGER Act). Since reporting is mandatory, the CDP offers a way for firms to publicize these efforts to a wider audience in reducing GHG emissions and signal to the public carbon-mitigation efforts to improve their reputation. Furthermore, although the effects of climate change are geographically variable, Australia has experienced both visible and widespread climate-induced disasters over the last decade and extreme events like flooding and bush-fires have become more frequent. Yet the country often ranks very low in comparable climate change performance indexes, ranking last behind 193 United Nations member countries in the 2021 Sustainable Development Report. The domestic economy is heavily dependent on coal for electricity generation, as well as a major export commodity, and as such Australians are "high per capita contributors to anthropogenic climate change" (Head et al. 2014). These characteristics of a generally poor-climate-performing business environment, mandatory reporting, and tangible negative climate change-induced events, make Australia a favorable setting to test the effects of reputational pressure on firm behavior.

## 5.2 Data

Using Rvest, I scrape data from the Carbon Disclosure Project on Australian firms between 2010-2021, with yearly reports on which firms respond to the CDP and their corresponding grade and scoring. In total there are 383 unique firms, with observations covering a twelve year period, amounting to 4,888 observations. Due to missing data, some observations are dropped from the analysis. There are two dependent variables; scoring and response. Responded is a binary variable that takes the value of 1 if the firm responded for evaluation that year, and 0 otherwise. Since Responded also captures firms who 'fail to respond' I include an additional variable 'Responded_Robust' and drop these observations. Firms are scored between 0-100 and assigned a grade between A-F based on the firm's progress towards "environmental stewardship". Each scoring is sector-specific, allowing a peer-to-peer benchmarking. I recode the scoring variable to be a weighted variable between 0 and 1; companies scoring A grades closer to 1 while companies scoring F grades closer to 0. The global stock market return data is from the Thomson Refinitiv datastream database. I compile firm level data from Refinitiv Eikon, and use a fuzzy matching command to match company stock-market codes. I cross-checked firm matches. Since the data is unbalanced, I impute missing values for continuous (binary) variables based on linear (logistic) regression models. Descriptive statistics of the distribution of sectors is shown in figure 2.

## 5.3 Explanatory variables

There is a dichotomy between B2B (business to business) firms and B2C (business to consumer) firms (Lilien 2016). In B2B markets, organizational buyers will purchase goods and services on behalf of businesses, organizations or insti-

tutions. In contrast, in consumer markets, individual consumers purchase goods and services for their own consumption and that of other individual consumers. Many past accounts on corporate social performance have omitted PMP from theoretical or empirical models,[7] since these studies have used datasets conflating B2B and B2C firms, the heterogeneous effects of firm audiences to firm action has not been fully addressed. B2C firms face a larger 'public' audience, while B2B firms are shielded from consumer audiences, and instead face businesses. Thus, public and consumer attention differs between B2B and B2C firms and therefore reputational pressure differs widely between the two.

My independent variable is PMP (product-market-profile) which reflects the institutional structure of the firm and their sensitivity to consumers versus business audiences. To construct the variable, I follow recent research (Srinivasan, Lilien, and Sridhar 2011; Luffarelli et al. 2019), and use a well-established classification based on four-digit SIC codes.[8] As Refinitiv only use NAICS codes, I manually recode six-digit NAICS codes to SIC codes using the `Concordance` package in R. PMP is a binary indicator variable that equals 1 if a firm principally operates in business markets and 0 if the firm principally operates in consumer markets. This classification is based on the idea of product-market profiles reflect firms operating in either B2B markets or B2C markets (Dwyer and Tanner 2002), however a secondary scheme of product-market profiles conceives of firms as either offering goods or services (Parasuraman, Zeithaml, and Berry 1988).[9]

[7] For exceptions see Flammer (2015), M. Johnson, Redlbacher, and Schaltegger (2018), and Servaes and Tamayo (2013).

[8] Firms with the following four-digit SIC codes are categorized as B2C: 2011–2273, 2299–2399, 2511–2519, 2676–2741, 2771, 2842, 2844, 2851, 3142–3199, 3261–3269, 3421, 3634–3641, 3711, 3714–3716, 3751, 3911, 3931–3952, 3961–3991, 3995, 4121, 4724, 4725, 5012, 5021, 5023, 5091, 5093–5099, 5122–5154, 5181, 5182, 5192–6011, 6411–6515, 6531–6541, 7011–7299, 7513–8699, and 8811. Firms with the other four-digit SIC codes are categorized as B2B

[9] It is possible that B2C firms offering goods may face varied reputational pressure than

I use the carbon emissions measurements created by Cory, Lerner, and Osgood (2021) to capture industries that both directly emit carbon dioxide, heavily consume CO2-intensive electricity, and suppliers of carbon-intensive industries. Although climate policy and climate-orientated activists focus on first-in-line industries (direct emitters and electricity consumers), carbon intensity affects the entire industrial supply chain, affecting a greater set of firm-level actors. I use estimates on CO2 carbon emissions at the industry-level (Henry, Khan, and Cooke-Hull 2010). 'Direct' carbon emissions is a single variable that refers to industries that either emit CO2 through their own activities or consume electricity that emits CO2. This measurement satisfies the literature's findings on both areas of importance for quantifying emissions. Australia's industrial sectors vary widely in the intensity and degree of their carbon emissions, which is the ratio of emissions generated to the value of the output produced by a given industry or sector.

I use two measures of exposure to a climate-disaster. I first use data on climate-induced natural disasters from the EMDAT dataset in Australia spanning 2010-2021. I cross-check all disasters that are specifically linked to climate change, and aggregate the number of climate disasters experienced each year (count_dis). I include variables related to magnitude of the disaster, number of deaths, and number of people affected.

Additionally, to capture the real impact on consumer audiences of natural disasters that triggers climate awareness I use a proxy for the potential total

those offering services, since there may be a stronger relationship between employees and customers in the latter. B2B and B2C firms could be disaggregated further into four categories to account for the two variations in product-market profiles (B2C goods, B2C services, B2B goods, and B2C services).

economic losses from natural disasters linked to climate disasters by using household insured losses from the Insurance Council Australia database. I measure aggregate amounts of total claims received, original loss value, and normalized loss value. There are some limitations to using this proxy. Insurance data tallies insured losses only, and as such accounts for only a proportion of the total cost of a disaster. Insurance coverage is patchy - with residential buildings underinsured and insurance penetration coverage of crops and fences limited (Keating and Handmer 2011). A number of intangible costs are difficult to capture with insurance data, for example (non) residential clean-up, disruption or loss of public services, and disaster response and relief (Ladds et al. 2017). Despite these limitations, insurance data is a useful standardized proxy, and where possible I include alternative measurements of natural disaster effects from EMDAT.

## 5.4 Controls

An additional consideration is that my dataset of publicly traded companies contains firms of all sizes. Particularly when market-capitalisation weight is used, larger companies can disproportionately dominate the results, since larger companies have both the time and resources for carefully-crafted responses, as well as increasing their public profile and costs of non-compliance/poor behavior. I therefore include a control for company size. Following standardized practice, I define small cap companies as any company with a market capitalization up to $2 billion; mid-cap companies as any company with a market capitalisation between $2 billion and $10 billion; and large cap as any company with a market capitalisation over $10 billion. Market cap are measured in USD. I also use company size to determine whether large 'early movers' can induce reputational effects between smaller firms.

Past studies have used ESG as a measurement of a firm's socially-consious reputation. To proxy for a firm's social consciousness or social capital (Guiso, Sapienza, and Zingales 2008a; Guiso, Sapienza, and Zingales 2008b; Guiso, Sapienza, and Zingales 2009), I use the firm's ESG score derived from Refinitiv Thomas Eikon. The literature has mixed findings on whether firms with high ESG scores are "followers" or "leaders" (Dyck, Volchkova, and Zingales 2008). While there is much evidence for the latter, the former represents the "dark view" of firms with high ESG scores. For instance, Cohen, Gurun, and Nguyen (2020) show that green-tech innovation is actually driven by oil and gas firms instead of firms with high environmental scores. I include data on industry level classification from Refinitiv Eikon.

# 6 Results

The main statistical analysis consists of a set of 2FE models estimated by ordinary least squares (OLS), probit and fractional logit models - adjusted to the dependent variable. The four general estimating equations are:

$$Y_{it} = \gamma_i + \lambda_t + \delta B2C_{it} + X_{it}\beta + \epsilon_{it} \tag{1}$$

$$Disclosure_{it} = \gamma_i + \lambda_t + \delta B2C_{it} + X_{it}\beta + \epsilon_{it} \tag{2}$$

$$Y_{it} = \gamma_i + \lambda_t + \delta B2C_{it} * \delta GHGintensity_{it} + X_{it}\beta + \epsilon_{it} \tag{3}$$

$$Scoring_{it} = \gamma_i + \lambda_t + \delta B2C_{it} * \delta GHGintensity_{it} + X_{it}\beta + \epsilon_{it} \qquad (4)$$

In equation (1), $Y_{it}$ is firm $i$'s decision to disclose information in year $t$, and is equal to either 0 or 1. In equation (2) and (4) $Y_{it}$ is a firm's scoring on internal climate regulations and policies. $\gamma_i$ represents a set of firm (and in some models, industry-level) fixed effects. $\gamma_i$ is a set of year fixed-effects. $\delta B2C$ is a dummy variable that indicates whether the firm is B2C or B2B. For equation (3) and (4) B2C is interacted with $\delta GHGintensity$ which is the intensity of emissions in the sector. $\lambda_t$ is the parameter of primary interest, the effect of B2C (and emission intensity) on firm behavior.$X_{it}$ is a matrix of covariates, $\beta$ a vector of coefficients, $\epsilon_{it}$ is the error term. The two-way fixed effects model controls for unobserved variation in the firm, sector or year levels. Without making causal claims, these results are interpreted as possible relationships between institutional structure and emissions intensity of firms and their behavior in disclosing information and performance on climate change.

Table 1 displays the results of four models. The point estimates are standard OLS coefficients, with robust standard errors clustered at the firm and year. As predicted, the B2C status variable is negative and significant across different specifications. To ensure the significance of the variable is not an artifact of the selected covariates, I report the bivariate model. The variable remains positive, and indeed increases in effect when adding firm and year fixed effects in models 2, 3, and 4. The control variables on ESG scoring and direct emissions, and company-market capitalisation and firm size are understandably significant. ESG score, company market capitalisation, and company size are

positively associated with disclosure, while high emitting firms are negatively associated with disclosure. The added controls do not absorb the variation in the dependent variable, and it remains substantive: the most conservative estimate is that B2C firms are 9.7 percentage points less likely to disclose to the CDP than B2B firms. Model (4) with all controls and fixed-effects estimates that B2C firms are 27.7 percent less likely to disclose information to the CDP than B2B firms.

In the Appendix I add three way fixed effects - firm, industry and year - not only do the results hold, but the substantive effect increases greatly. Firm fixed effects control for unobservable, time-invariant characteristics of individual companies that might influence their disclosure behavior, such as corporate culture or governance structures. By including year fixed effects, I account for global or national trends over time, such as evolving regulatory environments or shifting public awareness about climate issues, which could uniformly affect firms' propensity to disclose emissions. Industry fixed-effects are crucial, as they allow me to control for industry-specific factors—such as regulatory scrutiny or consumer pressure—that might systematically influence disclosure practices within certain sectors.

Table 1: OLS Regression Results: B2C Firms on CDP Disclosures

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Disclosure Dummy | | | |
| | *OLS* | | *probit* | |
| | (1) | (2) | (3) | (4) |
| Business to Consumer Dummy | $-0.070^{***}$ (0.022) | $-0.097^{**}$ (0.041) | $-0.187^{***}$ (0.060) | $-0.277^{**}$ (0.138) |
| ESG Score | | $0.007^{***}$ (0.001) | | $0.020^{***}$ (0.002) |
| Direct Emissions | | $-0.001^{***}$ (0.0005) | | $-0.003^{**}$ (0.001) |
| Company Market Cap (USD) | | $0.000^{***}$ (0.000) | | $0.000^{***}$ (0.000) |
| NAICS Admin/Waste | | 0.495 (0.319) | | 5.261 (265.347) |
| NAICS Agriculture | | 0.162 (0.324) | | 0.418 (277.752) |
| NAICS Arts/Rec | | 0.479 (0.316) | | 5.228 (265.347) |
| NAICS Construction | | $0.514^{*}$ (0.310) | | 5.186 (265.347) |
| NAICS Education | | 0.559 (0.375) | | 5.400 (265.348) |
| NAICS Finance/Insurance | | $0.598^{*}$ (0.306) | | 5.479 (265.347) |
| NAICS Health | | 0.217 (0.313) | | 3.361 (265.347) |
| NAICS Information | | 0.507 (0.335) | | 5.265 (265.347) |
| NAICS Manufacturing | | $0.518^{*}$ (0.307) | | 5.262 (265.347) |
| NAICS Mining/Energy | | $0.637^{**}$ (0.306) | | 5.568 (265.347) |
| NAICS Services | | $0.640^{*}$ (0.336) | | 5.698 (265.347) |
| NAICS Science/Tech | | 0.308 (0.314) | | 4.643 (265.347) |
| NAICS Real Estate | | $0.695^{**}$ (0.308) | | 5.768 (265.347) |
| NAICS Retail Trade | | $0.783^{**}$ (0.318) | | 5.854 (265.347) |
| NAICS Transport/Warehouse | | 0.478 (0.309) | | 5.088 (265.347) |
| NAICS Utilities | | $0.624^{*}$ (0.319) | | 5.590 (265.347) |
| NAICS Wholesale Trade | | 0.453 (0.310) | | 5.012 (265.347) |
| Company Size: Medium | | $-0.040$ (0.043) | | $-0.064$ (0.141) |
| Company Size: Large | | $-0.203^{***}$ (0.048) | | $-0.583^{***}$ (0.159) |
| Constant | $0.393^{***}$ (0.011) | $-0.225$ (0.312) | $-0.272^{***}$ (0.029) | $-5.974$ (265.347) |
| Observations | 2,663 | 1,377 | 2,663 | 1,377 |
| $R^2$ | 0.004 | 0.265 | | |
| Adjusted $R^2$ | 0.003 | 0.252 | | |
| Akaike Inf. Crit. | | | 3,514.033 | 1,493.216 |

*Note:* $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

**B2C Firms and CDP Scoring: 2FE**

| | *Dependent variable:* | | |
|---|---|---|---|
| | CDP Scoring | | |
| | *OLS* | *panel linear* | |
| | (1) | (2) | (3) |
| Business to Consumer | 0.025** (0.012) | 0.205 (0.142) | 0.063 (0.040) |
| Emission Intensity | | 0.002*** (0.001) | 0.005*** (0.0005) |
| Direct Emissions | | -0.003** (0.001) | -0.001* (0.0003) |
| Company Size: Medium | | -0.044 (0.028) | -0.040** (0.020) |
| Company Size: Small | | -0.024 (0.033) | -0.064*** (0.023) |
| NAICS Sector (Administrative and Support and Waste Management and Remediation Services) | | | 0.106 (0.160) |
| NAICS Sector(Agriculture, Forestry, Fishing and Hunting) | | | 0.008 (0.162) |
| NAICS Sector (Arts, Entertainment, and Recreation) | | | -0.019 (0.159) |
| NAICS Sector (Construction) | | | 0.185 (0.155) |
| NAICS Sector (Educational Services) | | | 0.081 (0.187) |
| NAICS Sector (Finance and Insurance ) | | | 0.117 (0.152) |
| NAICS Sector (Health Care and Social Assistance) | | | 0.001 (0.157) |
| NAICS Sector (Information) | | | -0.051 (0.171) |
| NAICS Sector (Manufacturing) | | | 0.052 (0.152) |
| NAICS Sector (Mining, Quarrying, and Oil and Gas Extraction) | | | 0.030 (0.153) |
| NAICS Sector (Other Services (except Public Administration)) | | | 0.013 (0.172) |
| NAICS Sector (Professional, Scientific, and Technical Services) | | | 0.027 (0.158) |
| NAICS Sector (Real Estate and Rental and Leasing) | | | 0.223 (0.154) |
| NAICS Sector (Retail Trade) | | | 0.203 (0.163) |
| NAICS Sector (Transportation and Warehousing) | | | 0.109 (0.155) |
| NAICS Sector (Utilities) | | | 0.213 (0.162) |
| NAICS Sector (Wholesale Trade) | | | 0.121 (0.155) |
| B2C*ESG | | | -0.001* (0.001) |
| Constant | 0.235*** (0.006) | | |
| Observations | 2,663 | 1,377 | 1,377 |
| $R^2$ | 0.002 | 0.021 | 0.285 |
| Adjusted $R^2$ | 0.001 | -0.125 | 0.121 |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 |

Table 2

Table 2 reports the results using CDP scoring as the dependent variable. The point estimates are standard OLS coefficients, with robust standard errors clustered at the firm and year. Although the B2C variable is positive and significant in the bivariate model, adding fixed-effects and additional covariates largely absorbs this effect, and B2C has no significant effect on the scoring results of firms that do submit to the CDP. To test whether firms who are generally high ESG performers and also B2C facing would score higher, the results suggest this

is not the case. Although substantively small, B2C firms with high ESG scores are slightly more likely to score worse on CDP climate reports. Interestingly, though table 1 suggests that high emitting firms are slightly less likely to disclose information to the CDP, the high-emitting firms that do submit score higher than low-emitting firms. This would not be incompatible with the idea that high-emitting firms only submit if they are certain they would receive a positive rating, and make reputational gains by signaling their climate-consciousness. High-emitting firms with poor internal regulations would avoid disclosing and suffering from even greater public scrutiny. Table 3 (see appendix) largely confirms this hypothesis, where the interaction effect of B2C and high-intensity emitters results in a 15.5 percent increase in the liklihood these firms disclose information to the CDP.

## 6.1 Natural Disasters

The second statistical analysis consists of 2FE models estimated by OLS models. These models swap out B2C firms for several variables capturing high-impact climate-linked natural disasters. In order to measure the effects of climate-linked disasters on firm disclosure and climate performance, it is critical to confront issues of selection. Although I control for confounders and time-variant effects, firms who disclose may be systematically different than non-disclosures on some observable metric (e.g. latent value of disclosing on firm value) resulting in biased inferences. To further account for potential endogeneity issues I employ a stacked difference-in-difference approach. I use a multiple time periods DiD design that adopts a staggered treatment assumption, and parallel trends assumption based on not-yet treated units taking the group-time average treatment effect. This approach can help address factors that change over time and

could influence both the likelihood of experiencing natural disasters and the decision to disclose emissions, for instance there could be some confounding variable relationship with annual increases in temperature or relative humidity increases and climate events and decisions to disclose or not disclose. Even though I do control for ESG, firms with a stronger commitment to environmental practices that is not captured by ESG might be more likely to disclose, regardless of experiencing disasters. Lastly, if certain economic factors both influence a firm's liklihood of experiencing natural disasters, such as manufacturing or distribution hubs being located in vulnerable areas, and its decisions to disclose, failing to account for this could skew the results. The proposed linear hierarchical model for the change in disclosure levels before and after years with high-impact natural disasters between firms i and j is given by:

$$\sum[\Delta\tau_{ijth}\chi_{ijt}^{high-disaster}]-\sum[\Delta\tau_{ijth}\chi_{ijt}^{lowdisaster}] = \beta^{highdisaster}+\gamma_{h}^{lowdisaster} \quad (5)$$

Here $\Delta\tau_{i}jth$ represents a change in the average likelihood of disclosure i and $j$ at time t. $\chi_{ijt}^{high-disaster}$ is an indicator of whether the year has experienced acute natural disasters linked to climate-change. For additional robustness, I use three measurements of high-disaster treatment. First, I code the number of climate-linked natural disasters that have occurred in a given year; three or more disasters constitutes a 'high' disaster year. I then use the total number of insured damages that year, exploiting a sizeable jump in the data with four years of high-insured costs of climate disasters. I then use the total number of individuals affected by climate disasters in a given year, again exploiting a

sizeable jump between high and low numbers affected. All three measurements capture different dimensions of the material, tangible costs of climate disasters on individuals that might be missed by a single measure. I take the lead variable of disclosure status as the outcome. Table 3 depicts the results of the estimator, showing a negative and statistically significant relationship between high-disasters and subsequent submission. As predicted, the most conservative estimate from the measurements (total number of individuals affected) predicts firms are 15.9 percent less likely to disclose to the CDP following a high-disaster year. The most favourable estimate (number of natural disasters) predicts a 22.3 percent decline in disclosures following high-disaster years. As a further robustness and sensitivity check, I recode the high-disaster variable to include 4 or more disasters. The results remain significant. The second part of my hypothesis relies on this relationship 'switching' for firms that do submit following high disaster years. Table 4 depicts a positive and significant relationship between disasters and CDP scoring. Firms that do submit to the CDP following high-disaster years are more likely to receive higher scores. These findings suggest that when Australia experiences a high-number of acute climate-linked natural disasters with tangible negative effects on individuals, firms are much less likely to disclose to the CDP. As a signaling mechanism, disclosure can bring unwanted public scrutiny if the firm has a poor climate track-record. However, firms that do disclose information, score much higher in the years that follow high climate-linked natural disasters. Since audiences are primed to care about a firm's climate-related behavior, these firms use this as an opportunity to rebuild or repair their reputation.

**Difference-in-Difference OLS models**

| | *Dependent variable:* | | |
|---|---|---|---|
| | Company Score | | |
| | (1) | (2) | (3) |
| treated_Dis | -0.034*** (0.013) | | |
| time1 | -0.242*** (0.009) | | |
| did_dis | 0.201*** (0.016) | | |
| treated_TA | | -0.066*** (0.010) | |
| time2 | | -0.207*** (0.008) | |
| did_affected | | 0.139*** (0.014) | |
| treated_insured | | | -0.040*** (0.013) |
| time3 | | | -0.248*** (0.009) |
| did_insured | | | 0.168*** (0.016) |
| Constant | 0.266*** (0.005) | 0.318*** (0.007) | 0.272*** (0.004) |
| Observations | 4,808 | 4,808 | 4,808 |
| $R^2$ | 0.131 | 0.122 | 0.152 |
| Adjusted $R^2$ | 0.130 | 0.122 | 0.152 |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 |

Table 3

One key limitation to the results would be if there is a dynamic treatment effect, where the impact of experiencing a natural disaster on decisions to disclose depends on past exposure to disasters. This stacked difference-in-difference model may not be able to account for that. As this study is limited in both the sample period, the annual coverage, and not using geo-locations for headquarters, manufacturing and distribution centres, testing for a dynamic treatment would be difficult. Future research could uncover this with a longer time-series dataset for countries with longer histories of company disclosures.

It might be the case that firms do not find negative reporting costly, nor may the prospect of negative reporting impact their behavior in terms of disclosure.

**Difference-in-Difference OLS models**

| | *Dependent variable:* | | |
|---|---|---|---|
| | Disclosure Dummy:lead | Disclosure Dummy:lead | |
| | (1) | (2) | (3) |
| treated_Dis | 0.089*** (0.024) | | |
| time1 | 0.168*** (0.017) | | |
| did_dis | -0.223*** (0.030) | | |
| treated_TA | | 0.028 (0.018) | |
| time2 | | 0.103*** (0.016) | |
| did_affected | | -0.159*** (0.025) | |
| treated_insured | | | 0.102*** (0.024) |
| time3 | | | 0.151*** (0.016) |
| did_insured | | | -0.218*** (0.031) |
| Constant | 0.596*** (0.009) | 0.600*** (0.012) | 0.600*** (0.008) |
| Observations | 4,809 | 4,809 | 4,809 |
| $R^2$ | 0.022 | 0.016 | 0.019 |
| Adjusted $R^2$ | 0.021 | 0.015 | 0.018 |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 |

Table 4

I therefore test whether negative reporting by the CDP impacts subsequent firm behavior. I use a 2FE models estimated by ordinary least squares (OLS), and test the effect of a failing or low grade ('D' or 'F') from firms in a given reporting year. The results are depicted in table 5 showing that firms that receive a failing grade are 31.1 percent less likely to resubmit to the CDP in the years that follow. This suggests that reports on poor climate performance from the CDP may be costly for the firm. I also include an interaction effect with B2C facing firms; the effect is both significantly negative and substantial. This lends further support to the hypothesis that B2C facing firms are deterred from disclosing to avoid public scrutiny if their poor performance on climate is exposed. Since these firms face higher reputational costs than B2B firms, they are even

less likely to disclose information to the CDP following a failing grade.

**Effect of Failing Grade on Future Disclosure (Year & Firm Two-Way Fixed-Effects)**

| | *Dependent variable:* | |
|---|---|---|
| | Disclosure Dummy | |
| | (1) | (2) |
| failing_grade_dummy | -0.364*** (0.027) | -0.314*** (0.032) |
| ESG | 0.003*** (0.001) | 0.003*** (0.001) |
| company_sizemedium | | -0.146*** (0.049) |
| company_sizesmall | | -0.186*** (0.058) |
| failing_grade_dummy:B2C_Dumy | | -0.137** (0.053) |
| B2C_Dumy | -0.929*** (0.245) | -0.782*** (0.250) |
| Observations | 1,377 | 1,377 |
| $R^2$ | 0.156 | 0.169 |
| Adjusted $R^2$ | 0.032 | 0.044 |
| *Note:* | | *p<0.1; **p<0.05; ***p<0.01 |

Table 5

Company decisions to disclose could be attributed to a cultural or normative practice within the company. It makes sense that as companies begin to submit to the CDP this process is then integrated into the companies' annual responsibilities, and actively implement into job descriptions to ensure a report is submitted to the CDP. This would suggest that even when a company receives a poor grade, the company will continue to submit regardless as this practice has been integrated into the normative culture. One case study of a demerger of Australian mining companies illustrates why decisions to disclose are less related to cultural norms within the company, and more related to reputational payoffs. BHP Billiton is one of Australia's largest mining conglomerates. The company ranks 20th as of 2020-2021 in total emissions, clocking in 3,742,988 tons of scope 1 CO2 emissions. In 2015 the company demerged and created a second company, South32, to divest its second-tier assets, based in Western Australia. Despite the fracturing, BHP staff were retained and carried on their

roles at the new South32, including the majority of the executive board.

Prior to the de-merger, BHP had been submitting reports to the CDP for disclosure to receive scoring since 2010, scoring an average B grade with a smattering of A's and one C. In other words, BHP Billiton scored highly and performed extremely well in climate change reporting, particularly given its status as a high emitting mining company. The company has continued this practice every year for 14 years as of 2023. In contrast, the newly spun out company South32, continued the practice from its inaugural year of 2016 until 2018, when the company received a D scoring through disclosure. After this the company declined to participate for every subsequent year. The company culture at South32 and the in-built practice of disclosure was carried on into the new company; if there had been some kind of cultural shift then South32 would not have disclosed immediately, and if South32 simply didn't have the resources like BHP had to disclose, it also would have made sense to cease this practice. Instead, the decision to report to disclose and report to the CDP carried on for three years only stopping when the company was hit with a failing grade. The reputational gains of disclosure and reputational costs of being found with a poor grade help explain why South32 stopped disclosing. In this case, South32 is a B2B firm, and we might expect it to behave as other poor-scoring B2B firms do – that is, they submit for disclosure and are less concerned with poor grading since this is relatively obscured. In South32's case, they have the problem of co-comparison as the offspring of BHP, where their scores are directly compared to their parent company. The relative benchmark for South32 is not the rest of the industry, it's BHP and its outstanding performance on CDP reports.

Turning to reputational externalities (hypothesis 3) and reputational cascades (hypothesis 3). I again use 2FE models estimated by ordinary least squares (OLS). On reputational externalities, I find no support for within-industry effects. For reputational cascades, I regress large firms (coded from their company market capitalisation) on disclosure. Table 6 shows that conditional on a large-firm within the sector submitting, firms are more likely to disclose to the CDP.

**Effect of Early Movers on Sector-Level Disclosure (Year & Firm Two-Way Fixed-Effects)**

| | *Dependent variable:* | |
|---|---|---|
| | Disclosure Dummy | |
| | *OLS* | *panel linear* |
| | (1) | (2) |
| large_submitted | 0.371*** (0.049) | 0.469*** (0.043) |
| ESG | 0.004*** (0.001) | 0.003*** (0.001) |
| B2C_Dumy | -0.024 (0.037) | -0.367 (0.243) |
| naics_sector_name.xAdministrative and Support and Waste Management and Remediation Services | -0.061 (0.289) | |
| naics_sector_name.xAgriculture, Forestry, Fishing and Hunting | 0.002 (0.297) | |
| naics_sector_name.xArts, Entertainment, and Recreation | -0.034 (0.290) | |
| naics_sector_name.xConstruction | -0.159 (0.287) | |
| naics_sector_name.xEducational Services | 0.157 (0.335) | |
| naics_sector_name.xFinance and Insurance | -0.067 (0.284) | |
| naics_sector_name.xHealth Care and Social Assistance | -0.196 (0.289) | |
| naics_sector_name.xInformation | 0.166 (0.307) | |
| naics_sector_name.xManufacturing | -0.186 (0.283) | |
| naics_sector_name.xMining, Quarrying, and Oil and Gas Extraction | -0.106 (0.285) | |
| naics_sector_name.xOther Services (except Public Administration) | -0.151 (0.305) | |
| naics_sector_name.xProfessional, Scientific, and Technical Services | -0.040 (0.286) | |
| naics_sector_name.xReal Estate and Rental and Leasing | -0.144 (0.286) | |
| naics_sector_name.xRetail Trade | -0.141 (0.297) | |
| naics_sector_name.xTransportation and Warehousing | -0.173 (0.286) | |
| naics_sector_name.xUtilities | 0.166 (0.289) | |
| naics_sector_name.xWholesale Trade | -0.059 (0.287) | |
| company_market_cap_usd | -0.000 (0.000) | |
| Constant | 0.494* (0.285) | |
| Observations | 1,405 | 1,377 |
| $R^2$ | 0.178 | 0.107 |
| Adjusted $R^2$ | 0.166 | -0.025 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 | |

Table 6

# 7 Conclusion

Why do some firms implement internal carbon regulations? Why do some firms seek to be seen as climate-friendly? Climate change is a "global market failure" resulting from excessive greenhouse gas emissions. Firms constitute a large actor in excessive pollution because such pollution is directly costless - however these harms are then 'externalised' on a distant other (either future generations or different locations), who then face the brunt of climate impacts. For the most part hyperbolic discounting and diffuse actors insulates firms from audience awareness of firm behavior on climate change. However, whether firms feel pressure to cultivate climate-conscious reputations can vary depending on the institutional structure of the firm and their exposure to consumer audiences, the association of the firm's industry as a high-emitting 'culprit', and whether or not the consumer is 'primed' to be aware of the negative externalities of climate change.

Using Australia as a case study, I argue that B2C facing firms are more sensitive to public scrutiny, and therefore avoid sending any signals of their performance on climate change, to avoid negative exposure. A similar relationship holds for high-emitters. However, high-emitting B2C facing firms already face acute public scrutiny over their environmental impacts, with little to loose, these firms can use CDP disclosure and reporting to repair their reputation. I theorize that natural disasters - climate shocks - can deter firms from submitting for reporting, however firms that do submit score substantially higher following natural disaster years. Again, since audience attention towards climate (and culpable agents) is sharpened, firms face a trade-off in disclosing information. Firms that do submit reports only do so if they can ensure a positive outcome, since the reputational costs of poor performance are heightened.

I theorize two additional mechanisms through which concern for reputation can affect firm behavior - reputational cascades and reputational externalities. I find robust support for these theories - following large early-movers (large companies within a sector), firms within that industry are more likely to follow suit. Similarly, firms from sectors with collectively poor reputations are more likely to disclose to the CDP in an effort to repair their reputation-by-association.

My theory advances the idea that public-interest climate policies are not so 'Osonian' but may also be 'Stiglerian' by providing reputational benefits and competitive advantages to some industry groups, while harming others. My argument does not conflict with the idea that market forces have shifted the economic interests of many major industrial firms towards natural gas expansion and conservative support for carbon pricing, nor does it contradict the idea of strategic accommodation. In fact, reputation could be considered part of a looser interpretation of strategic accommodation. Following Levy and Newell's (2005) idea that corporate support for environmental policies is a means of maintaining corporate legitimacy, and that support of carbon pricing enhances the firm's responsible image. Highly visible corporations are open to public criticism, if they ignore calls for climate action they will incur significant reputational costs.

This paper does not establish the efficacy of these semi-private regulatory bodies on regulating firm behavior on climate change, however it does suggest the conditions under which firms will voluntary self-regulate by going above what they are legally obligated to do. The results speak to the relationship between the civil society, quasi-regulatory bodies and corporate responsiveness. In the absence of timely, comprehensive domestic litigation on climate change

where stakeholders cannot easily or directly challenge powerful firms through legal channels like regulatory bodies or domestic judiciaries, self-regulation of firms through semi-private regulatory bodies is an important channel of changing firm behavior. This paper suggests that if reputational pressure can fluctuate in conjunction with consumer awareness, increasing attention and improving transparency on firm behavior on climate change may induce positive outcomes.

Another key limitation of this paper is the geographically concentrated nature of businesses and climate effects, as well as public opinion data and attitudes following climate shocks. Future research may attempt more granular measures of geo-coding climate disasters to company headquarters or business outlets, and use survey data to determine whether extreme weather events increase consumer attention to firm behavior on climate change. I leave it up to future research to pursue this.

Although the empirical analysis here focuses solely on Australia, future research could extend the analysis to other countries over different time-periods. A crucial limitation of this paper is the lack of data on the consumer-side and incorporating survey data on changing consumer attitudes is a promising area of inquiry. Future work should investigate whether business coalitions affect individual firm behavior in semi-regulatory carbon mitigation schemes. Existing research shows these enduring and emerging coalitions have shaped climate policy through lobbying activity (Downie 2017; Kennard 2020; Kim, Urpelainen, and Yang 2016). Future work may fruitfully examine how firm's use coalition participation in conjunction with their involvement in private-regulatory initiatives.

# References


Aklin, Michaël and Matto Mildenberger (2020). "Prisoners of the wrong dilemma: why distributive conflict, not collective action, characterizes the politics of climate change". In: *Global Environmental Politics* 20.4, pp. 4–27.

Anderson, Lisa R and Charles A Holt (1997). "Information cascades in the laboratory". In: *The American economic review*, pp. 847–862.

Anton, Wilma Rose Q, George Deltas, and Madhu Khanna (2004). "Incentives for environmental self-regulation and implications for environmental performance". In: *Journal of environmental economics and management* 48.1, pp. 632–654.

Avant, Deborah D, Martha Finnemore, and Susan K Sell (2010). *Who governs the globe?* Vol. 114. Cambridge University Press.

Baccini, Leonardo and Lucas Leemann (2021). "Do natural disasters help the environment? How voters respond and what that means". In: *Political Science Research and Methods* 9.3, pp. 468–484.

Barnett, Michael L (2007). "Stakeholder influence capacity and the variability of financial returns to corporate social responsibility". In: *Academy of management review* 32.3, pp. 794–816.

Barrett, Scott and Robert Stavins (2003). "Increasing participation and compliance in international climate change agreements". In: *International Environmental Agreements* 3.4, pp. 349–376.

Bechtel, Michael M and Jens Hainmueller (2011). "How lasting is voter gratitude? An analysis of the short-and long-term electoral returns to beneficial policy". In: *American Journal of Political Science* 55.4, pp. 852–868.

Benjamin, Beth A and Joel M Podolny (1999). "Status, quality, and social order in the California wine industry". In: *Administrative science quarterly* 44.3, pp. 563–589.

Bertels, Stephanie and John Peloza (2008). "Running just to stand still? Managing CSR reputation in an era of ratcheting expectations". In: *Corporate Reputation Review* 11.1, pp. 56–72.

Boyd, Brian K, Donald D Bergh, and David J Ketchen Jr (2010). "Reconsidering the reputation—performance relationship: A resource-based view". In: *Journal of management* 36.3, pp. 588–609.

Brammer, Stephen J and Stephen Pavelin (2006). "Corporate reputation and social performance: The importance of fit". In: *Journal of management studies* 43.3, pp. 435–455.

Budolfson, Mark Bryant et al. (2012). "Collective action, climate change, and the ethical significance of futility". In: *Princeton University*.

Bumpus, Adam G (2015). "Firm responses to a carbon price: corporate decision making under British Columbia's carbon tax". In: *Climate policy* 15.4, pp. 475–493.

Cable, Daniel M and Mary E Graham (2000). "The determinants of job seekers' reputation perceptions". In: *Journal of organizational Behavior* 21.8, pp. 929–947.

Calvert, Randall L (1987). "Reputation and legislative leadership". In: *Public Choice* 55.1, pp. 81–119.

Cheon, Andrew and Johannes Urpelainen (2013). "How do competing interest groups influence environmental policy? The case of renewable electricity in industrialized democracies, 1989–2007". In: *Political Studies* 61.4, pp. 874–897.

Cho, Charles H, Dennis M Patten, and Robin W Roberts (2006). "Corporate political strategy: An examination of the relation between political expenditures, environmental performance, and environmental disclosure". In: *Journal of Business Ethics* 67.2, pp. 139–154.

CiteDrive, Inc (2017). *Report. Carbon risk: a burning issue, Chapter 3*. URL: `https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Economics/Carbonriskdisclosure45/Report`.

Cohen, Lauren, Umit G Gurun, and Quoc H Nguyen (2020). *The ESG-innovation disconnect: Evidence from green patenting*. Tech. rep. National Bureau of Economic Research.

Colgan, Jeff D, Jessica F Green, and Thomas N Hale (2021). "Asset revaluation and the existential politics of climate change". In: *International Organization* 75.2, pp. 586–610.

Cory, Jared, Michael Lerner, and Iain Osgood (2021). "Supply chain linkages and the extended carbon coalition". In: *American Journal of Political Science* 65.1, pp. 69–87.

Cullenward, Danny and David G Victor (2020). *Making climate policy work*. John Wiley & Sons.

Damert, Matthias and Rupert J Baumgartner (2018). "External pressures or internal governance–what determines the extent of corporate responses to climate change?" In: *Corporate social responsibility and environmental management* 25.4, pp. 473–488.

Dasgupta, Susmita, Hemamala Hettige, and David Wheeler (2000). "What improves environmental compliance? Evidence from Mexican industry". In: *Journal of Environmental Economics and Management* 39.1, pp. 39–66.

Deephouse, David L (2000). "Media reputation as a strategic resource: An integration of mass communication and resource-based theories". In: *Journal of management* 26.6, pp. 1091–1112.

Deephouse, David L and Suzanne M Carter (2005). "An examination of differences between organizational legitimacy and organizational reputation". In: *Journal of management Studies* 42.2, pp. 329–360.

Djourelova, Milena et al. (2023). "Experience, Narratives, and Climate Change Beliefs". In.

Downie, Christian (2017). "Fighting for King Coal's crown: Business actors in the US coal and utility industries". In: *Global Environmental Politics* 17.1, pp. 21–39.

Dwyer, F Robert and John F Tanner (2002). "Business marketing: Connecting strategy, relationships, and learning". In.

Dyck, Alexander, Natalya Volchkova, and Luigi Zingales (2008). "The corporate governance role of the media: Evidence from Russia". In: *The Journal of Finance* 63.3, pp. 1093–1135.

Engel, Kirsten H and Scott R Saleska (2005). "Subglobal regulation of the global commons: The case of climate change". In: *Ecology LQ* 32, p. 183.

Esen, Emel (2013). "The influence of corporate social responsibility (CSR) activities on building corporate reputation". In: *International business, sustainability and corporate social responsibility*. Emerald Group Publishing Limited.

Eskander, Shaikh MSU and Sam Fankhauser (2020). "Reduction in greenhouse gas emissions from national climate legislation". In: *Nature Climate Change* 10.8, pp. 750–756.

Evans, Robert and Timothy W Guinnane (2006). "Reputational Externality and Self-Regulation". In.

Flammer, Caroline (2015). "Does corporate social responsibility lead to superior financial performance? A regression discontinuity approach". In: *Management Science* 61.11, pp. 2549–2568.

Fombrun, Charles J (2005). "Corporate reputations as economic assets". In: *The Blackwell handbook of strategic management*, pp. 285–308.

Gasper, John T and Andrew Reeves (2011). "Make it rain? Retrospection and the attentive electorate in the context of natural disasters". In: *American Journal of Political Science* 55.2, pp. 340–355.

Gazmararian, Alexander F and Helen V Milner (2022). "Political Cleavages and Changing Exposure to Global Warming". In: *Available at SSRN 4194232*.

Genovese, Federica (2019). "Sectors, Pollution, and Trade: How Industrial Interests Shape Domestic Positions on Global Climate Agreements". In: *International Studies Quarterly* 63.4, pp. 819–836.

Goel, Ran (2004). "A bargain born of a paradox: The oil industry's role in american domestic and foreign policy". In: *New Political Economy* 9.4, pp. 467–492.

Goins, Sheila and Thomas S Gruca (2008). "Understanding competitive and contagion effects of layoff announcements". In: *Corporate Reputation Review* 11.1, pp. 12–34.

Graham, Mary E and Pratima Bansal (2007). "Consumers' willingness to pay for corporate reputation: the context of airline companies". In: *Corporate reputation review* 10.3, pp. 189–200.

Grenadier, Steven R, Andrey Malenko, and Ilya A Strebulaev (2014). "Investment Busts, Reputation, and the Temptation to Blend in with the Crowd". In: *Journal of Financial Economics* 111.1, pp. 137–157.

Griskevicius, Vladas, Joshua M Tybur, and Bram Van den Bergh (2010). "Going green to be seen: status, reputation, and conspicuous conservation." In: *Journal of personality and social psychology* 98.3, p. 392.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales (2008a). "Social capital as good culture". In: *Journal of the European economic Association* 6.2-3, pp. 295–320.

— (2008b). "Trusting the stock market". In: *the Journal of Finance* 63.6, pp. 2557–2600.

— (2009). "Cultural biases in economic exchange?" In: *The quarterly journal of economics* 124.3, pp. 1095–1131.

Hacker, Jacob S and Paul Pierson (2002). "Business power and social policy: Employers and the formation of the American welfare state". In: *Politics & Society* 30.2, pp. 277–325.

Haddock-Fraser, Janet and Iain Fraser (2008). "Assessing corporate environmental reporting motivations: differences between 'close-to-market'and 'business-to-business' companies". In: *Corporate Social Responsibility and Environmental Management* 15.3, pp. 140–155.

Haddock-Fraser, Janet Elaine and Marielle Tourelle (2010). "Corporate motivations for environmental sustainable development: exploring the role of consumers in stakeholder engagement". In: *Business Strategy and the Environment* 19.8, pp. 527–542.

Hale, Thomas (2011). "A climate coalition of the willing". In: *The Washington Quarterly* 34.1, pp. 89–101.

Hargrave, Timothy J (2008). "The Voluntary Environmentalists: Green Clubs, ISO 14001, and Voluntary Environmental Regulations". In: *Corporate Reputation Review* 11.1, pp. 109–111.

Hazlett, Chad and Matto Mildenberger (2020). "Wildfire exposure increases pro-environment voting within democratic but not republican areas". In: *American Political Science Review* 114.4, pp. 1359–1365.

Head, Lesley et al. (2014). "Climate change and Australia". In: *Wiley Interdisciplinary Reviews: Climate Change* 5.2, pp. 175–197.

Henry, David, Beethika Khan, and Sandra Cooke-Hull (2010). *US Carbon Dioxide Emissions and Intensities over Time: A Detailed Accounting of Industries, Government and Households*.

Hepburn, Cameron, Stephen Duncan, and Antonis Papachristodoulou (2010). "Behavioural economics, hyperbolic discounting and environmental policy". In: *Environmental and Resource Economics* 46.2, pp. 189–206.

Hoffman, Andrew J (2001). *From heresy to dogma: An institutional history of corporate environmentalism*. Stanford University Press.

Hollyer, James R, B Peter Rosendorff, and James Raymond Vreeland (2015). "Transparency, protest, and autocratic instability". In: *American Political Science Review* 109.4, pp. 764–784.

Holmberg, Kim and Iina Hellsten (2015). "Gender differences in the climate change communication on Twitter". In: *Internet research* 25.5, pp. 811–828.

Hsueh, Lily (2017). "What Explains Participation and Effort in Voluntary Climate Action by Businesses?: Evidence from the Carbon Disclosure Project". In: *Evidence from the Carbon Disclosure Project (March 21, 2017)*.

Hung, Angela A and Charles R Plott (2001). "Information cascades: Replication and an extension to majority rule and conformity-rewarding institutions". In: *American Economic Review* 91.5, pp. 1508–1520.

Jackson, Van (2016). *Rival reputations: Coercion and credibility in US-North Korea relations*. Cambridge University Press.

Jetten, Jolanda et al. (2021). "Responding to climate change disaster: The case of the 2019/2020 bushfires in Australia." In: *European Psychologist* 26.3, p. 161.

Johnson, Matthew, Friederike Redlbacher, and Stefan Schaltegger (2018). "Stakeholder engagement for corporate sustainability: A comparative analysis of B2C and B2B companies". In: *Corporate Social Responsibility and Environmental Management* 25.4, pp. 659–673.

Kahneman, Daniel, Olivier Sibony, and Cass R Sunstein (2021). *Noise: A flaw in human judgment*. Little, Brown.

Keating, Adriana and John Handmer (2011). "The cost of disasters in Australia and Victoria: No straightforward answers". In: *VCCCAR Project: Framing Adaptation in the Victorian Context, Working Paper* 3.

Kelsey, Nina (2018). "Industry type and environmental policy: Industry characteristics shape the potential for policymaking success in energy and the environment". In: *Business and Politics* 20.4, pp. 615–642.

Kennard, Amanda (2020). "The enemy of my enemy: When firms support climate change regulation". In: *International Organization* 74.2, pp. 187–221.

Keohane, Robert O (2015). "The global politics of climate change: Challenge for political science". In: *PS: Political Science & Politics* 48.1, pp. 19–26.

Keohane, Robert O and David G Victor (2016). "Cooperation and discord in global climate policy". In: *Nature Climate Change* 6.6, pp. 570–575.

Kim, Sung Eun, Johannes Urpelainen, and Joonseok Yang (2016). "Electric utilities and American climate policy: lobbying by expected winners and losers". In: *Journal of Public Policy* 36.2, pp. 251–275.

King, Andrew A and Michael J Lenox (2000). "Industry self-regulation without sanctions: The chemical industry's responsible care program". In: *Academy of management journal* 43.4, pp. 698–716.

Kirilenko, Andrei P, Tatiana Molodtsova, and Svetlana O Stepchenkova (2015). "People as sensors: Mass media and local temperature influence climate change discussion on Twitter". In: *Global Environmental Change* 30, pp. 92–100.

Kolk, Ans and David Levy (2001). "Winds of change:: corporate strategy, climate change and oil multinationals". In: *European Management Journal* 19.5, pp. 501–509.

Kolk, Ans and Jonatan Pinkse (2007). "Towards strategic stakeholder management? Integrating perspectives on sustainability challenges such as corporate responses to climate change". In: *Corporate Governance: The international journal of business in society*.

Kreps, Gary A (1984). "Sociological inquiry and disaster research". In: *Annual review of sociology*, pp. 309–330.

Kryvasheyeu, Yury et al. (2016). "Rapid assessment of disaster damage using social media activity". In: *Science advances* 2.3, e1500779.

Kuran, Timur (1997). *Private truths, public lies: The social consequences of preference falsification*. Harvard University Press.

Ladds, Monique et al. (2017). "How much do disasters cost? A comparison of disaster cost estimates in Australia". In: *International Journal of Disaster Risk Reduction* 21, pp. 419–429.

Levy, David L and Daniel Egan (1998). "Capital contests: National and transnational channels of corporate influence on the climate change negotiations". In: *Politics & Society* 26.3, pp. 337–361.

Levy, David L and Ans Kolk (2002). "Strategic responses to global climate change: Conflicting pressures on multinationals in the oil industry". In: *Business and Politics* 4.3, pp. 275–300.

Lilien, Gary L (2016). "The B2B knowledge gap". In: *International Journal of Research in Marketing* 33.3, pp. 543–556.

Love, E Geoffrey and Matthew Kraatz (2009). "Character, conformity, or the bottom line? How and why downsizing affected corporate reputation". In: *Academy of management Journal* 52.2, pp. 314–335.

Luffarelli, Jonathan et al. (2019). "The effect of corporate social performance on the financial performance of business-to-business and business-to-consumer firms". In: *Corporate Social Responsibility and Environmental Management* 26.6, pp. 1333–1350.

Markussen, Peter and Gert Tinggaard Svendsen (2005). "Industry lobbying and the political economy of GHG trade in the European Union". In: *Energy Policy* 33.2, pp. 245–255.

Marland, Alex and Brooks DeCillia (2020). "Reputation and brand management by political parties: Party vetting of election candidates in Canada". In: *Journal of Nonprofit & Public Sector Marketing* 32.4, pp. 342–363.

Martin, Nigel and John Rice (2010). "Analysing emission intensive firms as regulatory stakeholders: a role for adaptable business strategy". In: *Business Strategy and the Environment* 19.1, pp. 64–75.

Matisoff, Daniel C, Douglas S Noonan, and John J O'Brien (2013). "Convergence in environmental reporting: assessing the Carbon Disclosure Project". In: *Business Strategy and the Environment* 22.5, pp. 285–305.

McAllister, Jordan H and Afiq bin Oslan (2021). "Issue ownership and salience shocks: The electoral impact of Australian bushfires". In: *Electoral Studies* 74, p. 102389.

Meckling, Jonas (2011). *Carbon coalitions: Business, climate politics, and the rise of emissions trading*. MIT Press.

Meckling, Jonas (2015). "Oppose, support, or hedge? Distributional effects, regulatory pressure, and business strategy in environmental politics". In: *Global Environmental Politics* 15.2, pp. 19–37.

Newell, Peter and Matthew Paterson (1998). "A climate for business: global warming, the state and capital". In: *Review of International Political Economy* 5.4, pp. 679–703.

Nordhaus, William (2015). "Climate clubs: Overcoming free-riding in international climate policy". In: *American Economic Review* 105.4, pp. 1339–70.

Ostrom, Elinor (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge university press.

Parasuraman, Ananthanarayanan, Valarie A Zeithaml, and Leonard L Berry (1988). "Servqual: A multiple-item scale for measuring consumer perc". In: *Journal of retailing* 64.1, p. 12.

Pinkse, Jonatan and Ans Kolk (2012). "Addressing the climate change—sustainable development nexus: The role of multistakeholder partnerships". In: *Business & Society* 51.1, pp. 176–210.

Posner, Michael I, Mary K Rothbart, et al. (2007). "Research on attention networks as a model for the integration of psychological science". In: *Annual review of psychology* 58, p. 1.

Potoski, Matthew and Aseem Prakash (2005a). "Covenants with weak swords: ISO 14001 and facilities' environmental performance". In: *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management* 24.4, pp. 745–769.

— (2005b). "Green clubs and voluntary governance: ISO 14001 and firms' regulatory compliance". In: *American journal of political science* 49.2, pp. 235–248.

Prakash, Aseem and Matthew Potoski (2006). "Racing to the bottom? Trade, environmental governance, and ISO 14001". In: *American journal of political science* 50.2, pp. 350–364.

Pulver, Simone (2007). "Making sense of corporate environmentalism: An environmental contestation approach to analyzing the causes and consequences of the climate change policy split in the oil industry". In: *Organization & environment* 20.1, pp. 44–83.

Renn, Ortwin and Bernd Rohrmann (2000). *Cross-cultural risk perception: a survey of empirical studies*. Vol. 13. Springer Science & Business Media.

Servaes, Henri and Ane Tamayo (2013). "The impact of corporate social responsibility on firm value: The role of customer awareness". In: *Management science* 59.5, pp. 1045–1061.

Sethi, S Prakash, Terrence F Martell, and Mert Demir (2016). "Building corporate reputation through corporate social responsibility (CSR) reports:

The case of extractive industries". In: *Corporate Reputation Review* 19.3, pp. 219–243.

Shamsie, Jamal (2003). "The context of dominance: an industry-driven framework for exploiting reputation". In: *Strategic management journal* 24.3, pp. 199–215.

Sisco, Matthew R, Valentina Bosetti, and Elke U Weber (2017). "When do extreme weather events generate attention to climate change?" In: *Climatic change* 143, pp. 227–241.

Skjærseth, Jon Birger (2013). "Governance by EU emissions trading: resistance or innovation in the oil industry?" In: *International Environmental Agreements: Politics, Law and Economics* 13.1, pp. 31–48.

Skodvin, Tora and Jon Skjaerseth (2003). *Climate change and the oil industry: Common problem, different strategies*. Manchester University Press.

Srinivasan, Raji, Gary L Lilien, and Shrihari Sridhar (2011). "Should firms spend more on research and development and advertising during recessions?" In: *Journal of Marketing* 75.3, pp. 49–65.

Stafford, Edwin R and Cathy L Hartman (1996). "Green alliances: strategic relations between business and environmental groups". In: *Business Horizons* 39.2, pp. 50–60.

Standifird, Stephen S (2001). "Reputation and e-commerce: eBay auctions and the asymmetrical impact of positive and negative ratings". In: *Journal of management* 27.3, pp. 279–295.

Stavins, Robert N (2011). "The problem of the commons: still unsettled after 100 years". In: *American Economic Review* 101.1, pp. 81–108.

Steinzor, Rena I (1998). "Reinventing environmental regulation: the dangerous journey from command to self-control". In: *Harv. Envtl. L. Rev.* 22, p. 103.

Sunstein, Cass R (2005). *Why societies need dissent*. Vol. 9. Harvard University Press.

— (2006). "The availability heuristic, intuitive cost-benefit analysis, and climate change". In: *Climatic change* 77.1, pp. 195–210.

— (2019). *How change happens*. Mit Press.

Tirole, Jean (1996). "A theory of collective reputations (with applications to the persistence of corruption and to firm quality)". In: *The Review of Economic Studies* 63.1, pp. 1–22.

Turban, Daniel B and Daniel M Cable (2003). "Firm reputation and applicant pool characteristics". In: *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 24.6, pp. 733–751.

Vandenbergh, Michael P, Jim Rossi, and Ian Faucher (2020). "The gap-filling role of private environmental governance". In: *Virginia Environmental Law Journal* 38.1, pp. 1–57.

Victor, David G (2011). "Collective Action at Local and Global Scales: The Next Collaborative Agenda". In: *Available at SSRN 1910816*.

Victor, David G, Frank W Geels, and Simon Sharpe (2019). "Accelerating the low carbon transition". In: *The Case for Stronger, More Targeted and Coordinated International Action. Brookings Institution. Available online at: https://www. brookings. edu/wp-content/uploads/2019/12/Coordinatedactionreport. pdf (accessed January 24, 2021)*.

Vormedal, Irja, Lars H Gulbrandsen, and Jon Birger Skjærseth (2020). "Big oil and climate regulation: business as usual or a changing business?" In: *Global Environmental Politics* 20.4, pp. 143–166.

Weber, Elke U and Eric J Johnson (2009). "Mindful judgment and decision making". In: *Annual review of psychology* 60, p. 53.

Weber, Elke U and Paul C Stern (2011). "Public understanding of climate change in the United States." In: *American Psychologist* 66.4, p. 315.

Willinger, Marc and Anthony Ziegelmeyer (1998). "Are more informed agents able to shatter information cascades in the lab?" In: *The economics of networks: Interaction and behaviours*. Springer, pp. 291–305.

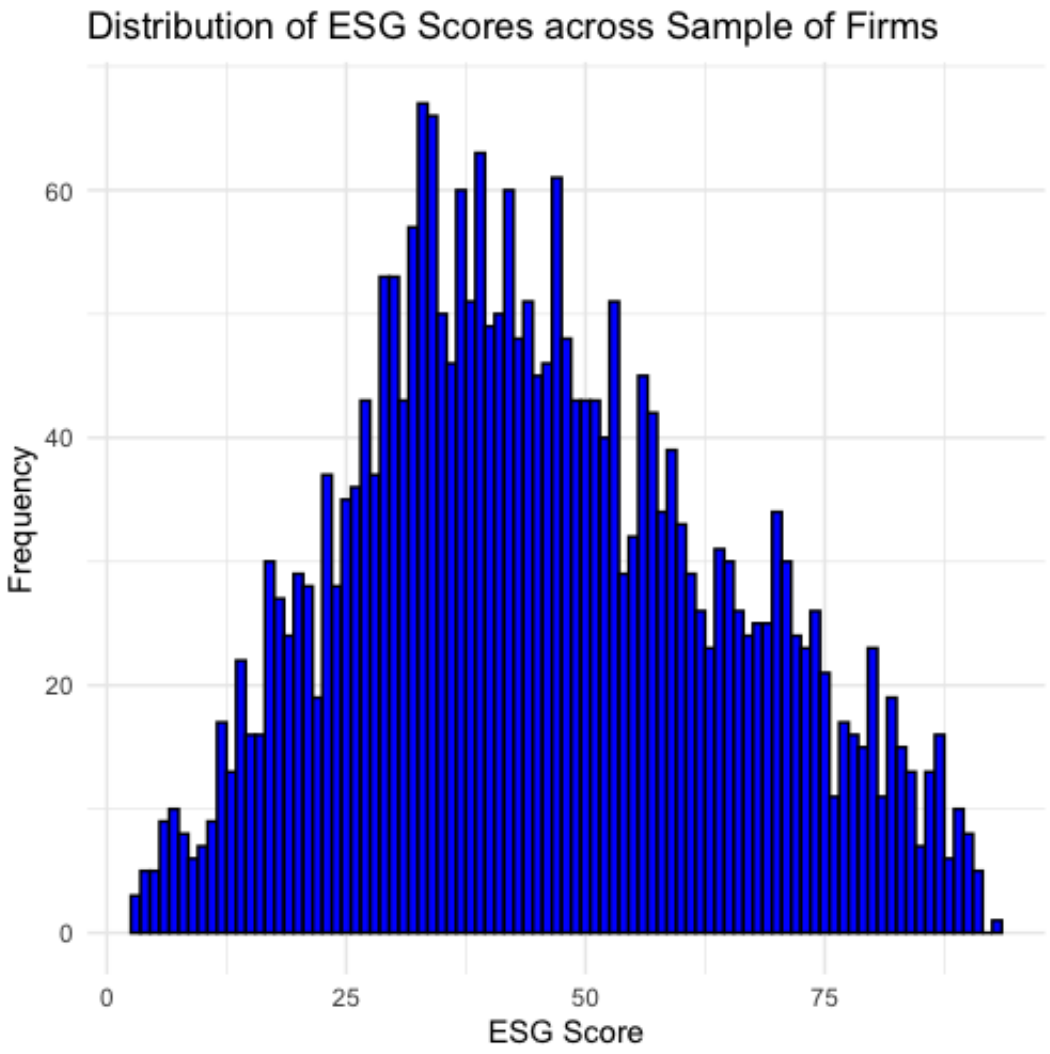
Winn, Monika I, Patricia MacDonald, and Charlene Zietsma (2008). "Managing industry reputation: The dynamic tension between collective and competitive reputation management strategies". In: *Corporate reputation review* 11.1, pp. 35–55.

Zhang, Yaming, Majed Abbas, and Wasim Iqbal (2022). "Perceptions of GHG emissions and renewable energy sources in Europe, Australia and the USA". In: *Environmental Science and Pollution Research* 29.4, pp. 5971–5987.

# A Appendix

## Descriptive Statistics

### Distribution of ESG Scores Across Sample



| Variable | Var | N | Mean | SD | Median | Trimmed |
|---|---|---|---|---|---|---|
| ESG | 1 | 2663.00 | 45.35 | 19.28 | 43.45 | 44.78 |
| NAICS Code* | 2 | 1377.00 | 7.84 | 4.14 | 7.00 | 7.64 |
| NAICS Sector* | 3 | 1377.00 | 10.48 | 4.06 | 10.00 | 10.50 |
| Large Firms | 4 | 2663.00 | 0.17 | 0.38 | 0.00 | 0.09 |
| Company Market Cap | 5 | 2663.00 | 7432460010.02 | 14580535491.56 | 2713448703.08 | 3906359173.99 |
| Disaster Magnitude Value | 6 | 2663.00 | 317025.90 | 640715.82 | 59000.00 | 152357.96 |
| Disclosure Response* | 7 | 2663.00 | 6.99 | 3.68 | 7.00 | 6.97 |
| CDP Status* | 8 | 2663.00 | 4.68 | 2.67 | 3.00 | 4.73 |
| Submission Dummy | 9 | 2663.00 | 0.37 | 0.48 | 0.00 | 0.34 |
| Weighted Score | 10 | 2663.00 | 0.24 | 0.25 | 0.25 | 0.20 |
| Grade | 11 | 2663.00 | 71.16 | 19.36 | 74.34 | 72.89 |
| environmental pillar ESG | 12 | 2663.00 | 47.74 | 22.82 | 46.85 | 47.91 |
| exec compensation: Environment | 13 | 2663.00 | 0.45 | 0.50 | 0.00 | 0.44 |
| executive members gender diversity % | 14 | 2663.00 | 55.65 | 23.86 | 58.45 | 56.27 |
| ESG Exec Compensation | 15 | 1373.00 | NA | NA | NA | NA |
| Firm Board Size | 16 | 2663.00 | 7.42 | 1.65 | 7.00 | 7.41 |
| Board Gender Diversity | 17 | 2663.00 | 32.22 | 10.88 | 33.33 | 32.65 |

Table 2: Descriptive Statistics - Part 1

| Variable | MAD | Min | Max | Range | Skew | Kurtosis | SE |
|---|---|---|---|---|---|---|---|
| ESG | 20.09 | 2.55 | 92.84 | 90.28 | 0.26 | -0.59 | 0.37 |
| NAICS Code* | 4.45 | 1.00 | 18.00 | 17.00 | 0.32 | -0.84 | 0.11 |
| NAICS Sector* | 4.45 | 1.00 | 18.00 | 17.00 | -0.03 | -0.55 | 0.11 |
| Large Firms | 0.00 | 0.00 | 1.00 | 1.00 | 1.76 | 1.11 | 0.01 |
| Company Market Cap | 3048400902.13 | 11417042.16 | 111797905561.50 | 111786488519.33 | 4.25 | 21.53 | 282545172.31 |
| Disaster Magnitude Value | 87473.40 | 0.00 | 2493931.00 | 2493931.00 | 2.66 | 6.21 | 12415.95 |
| Disclosure Response* | 4.45 | 1.00 | 13.00 | 12.00 | 0.04 | -1.18 | 0.07 |
| CDP Status* | 2.97 | 1.00 | 8.00 | 7.00 | 0.28 | -1.60 | 0.05 |
| Submission Dummy | 0.00 | 0.00 | 1.00 | 1.00 | 0.52 | -1.73 | 0.01 |
| Weighted Score | 0.19 | 0.00 | 1.00 | 1.00 | 1.47 | 1.73 | 0.00 |
| Grade | 20.87 | 0.00 | 100.00 | 100.00 | -0.75 | -0.03 | 0.38 |
| environmental pillar ESG | 24.77 | 0.00 | 95.78 | 95.78 | -0.01 | -0.70 | 0.44 |
| exec compensation: Environment | 0.00 | 0.00 | 1.00 | 1.00 | 0.20 | -1.96 | 0.01 |
| executive members gender diversity % | 25.89 | 14.94 | 99.75 | 84.81 | -0.21 | -1.02 | 0.46 |
| ESG Exec Compensation | NA | Inf | -Inf | -Inf | NA | NA | NA |
| Firm Board Size | 1.48 | 3.00 | 12.00 | 9.00 | 0.06 | 0.04 | 0.03 |
| Board Gender Diversity | 9.88 | 0.00 | 66.67 | 66.67 | -0.45 | 1.07 | 0.21 |

Table 3: Descriptive Statistics - Part 2

| Variable | Var | N | Mean | SD | Median | Trimmed |
|---|---|---|---|---|---|---|
| B2C Dummy | 1 | 2663.00 | 0.27 | 0.42 | 0.00 | 0.22 |
| Original Loss Value: Disaster | 2 | 2663.00 | 2237331667.78 | 1187319778.16 | 1681952650.00 | 2179616659.45 |
| Total Insurance Claims Received | 3 | 2663.00 | 70434.95 | 87668.45 | 45368.50 | 55195.49 |
| Commercial Property Claims | 4 | 2663.00 | 5144.68 | 5752.12 | 3578.00 | 4422.32 |
| Insurance Council Australia Metadata Rec | 5 | 2663.00 | 1995837.03 | 1453182.99 | 2218776.00 | 1945691.89 |
| Number Injured | 6 | 2663.00 | 16.04 | 44.26 | 0.00 | 2.99 |
| Number Homeless | 7 | 2663.00 | 639.21 | 1786.78 | 0.00 | 140.94 |
| Number Affected | 8 | 2663.00 | 21405.59 | 47621.88 | 600.00 | 8408.59 |
| Total Affected | 9 | 2663.00 | 22060.84 | 47504.33 | 820.00 | 9222.11 |
| Total Deaths from Disaster | 10 | 2663.00 | 9.39 | 12.76 | 3.00 | 7.36 |
| Wholesale Trade | 11 | 2663.00 | 1.00 | 0.00 | 1.00 | 1.00 |

Table 4: Descriptive Statistics - First Set of Variables

| Variable | MAD | Min | Max | Range | Skew | Kurtosis | SE |
|---|---|---|---|---|---|---|---|
| B2C Dummy | 0.00 | 0.00 | 1.00 | 1.00 | 1.02 | -0.82 | 0.01 |
| Original Loss Value: Disaster | 1273770030.10 | 409112234.00 | 4523469000.00 | 4114356766.00 | 0.47 | -0.83 | 23008172.20 |
| Total Insurance Claims Received | 67263.34 | 0.00 | 291074.00 | 291074.00 | 1.17 | 0.43 | 1698.86 |
| Commercial Property Claims | 5304.74 | 0.00 | 17334.00 | 17334.00 | 0.70 | -0.83 | 111.47 |
| Insurance Council Australia Metadata Rec | 1490546.74 | 0.00 | 4437238.00 | 4437238.00 | 0.07 | -1.10 | 28160.13 |
| Number Injured | 0.00 | 0.00 | 164.00 | 164.00 | 2.97 | 7.05 | 0.86 |
| Number Homeless | 0.00 | 0.00 | 6927.00 | 6927.00 | 3.20 | 8.35 | 34.62 |
| Number Affected | 889.56 | 0.00 | 186000.00 | 186000.00 | 2.83 | 6.87 | 922.83 |
| Total Affected | 1003.72 | 0.00 | 186000.00 | 186000.00 | 2.81 | 6.81 | 920.55 |
| Total Deaths from Disaster | 4.45 | 0.00 | 35.00 | 35.00 | 1.28 | -0.15 | 0.25 |
| Wholesale Trade | 0.00 | 1.00 | 1.00 | 0.00 | NA | NA | 0.00 |

Table 5: Descriptive Statistics - Second Set of Variables

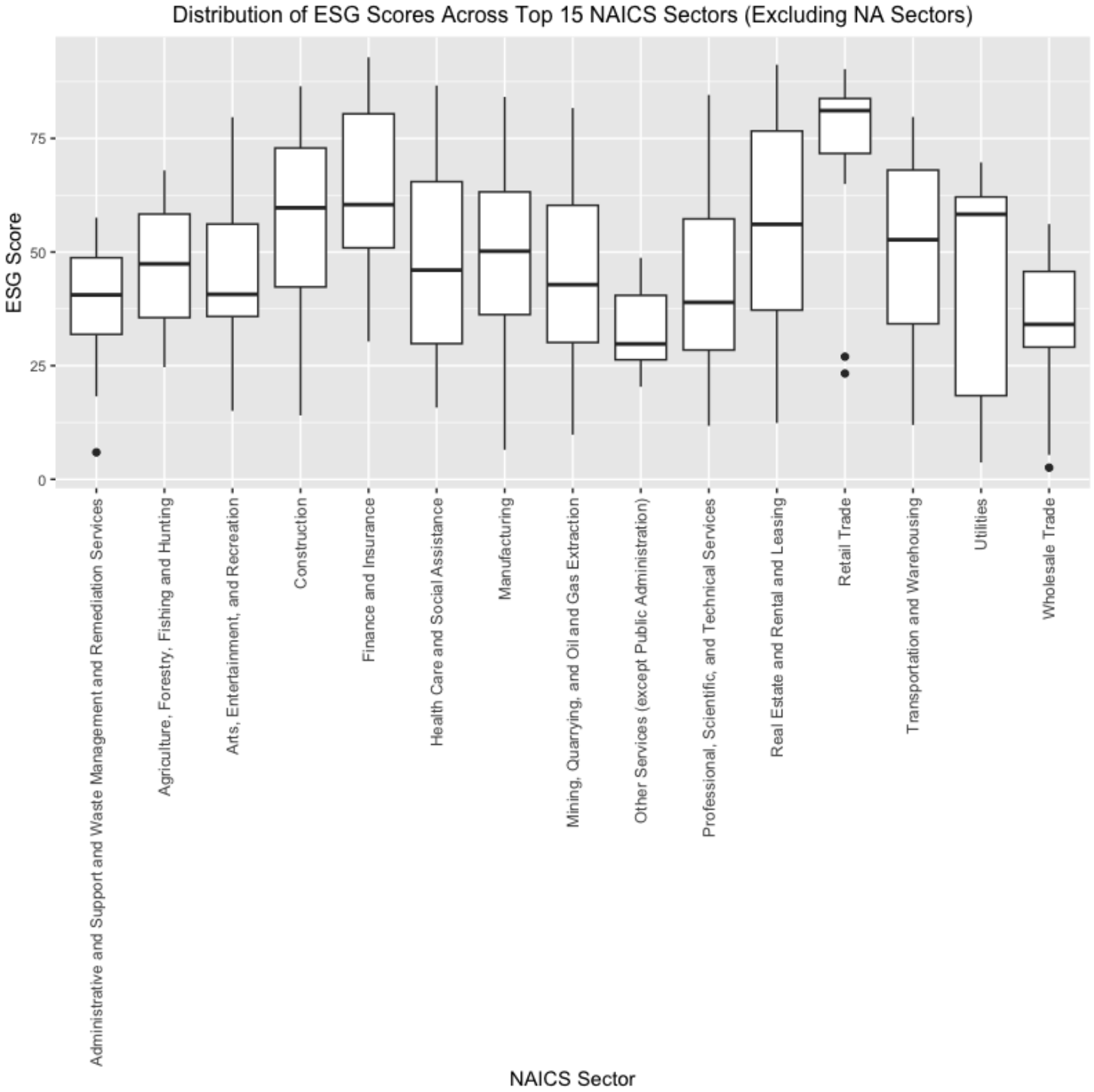



Figure 2: Distribution of ESG Scores Across Top 15 NAICS Sectors (Excluding NA Sectors)

# Data Collection Notes

# CDP Data

The CDP possesses the largest collection with global-coverage of foundational climate change, water and forest-risk information and utilizes these insights for strategic business decisions, investment approaches and policy choices.

## Description of Data Collection

I scraped data on companies, disclosures on emissions and climate-change behaviour and scoring by the CDP from their website; cdp.net. I used the cites 'Search and View past responses' and narrowed the search to companies in Australia, used yearly data from 2010 onwards (first year of CDP collecting data in Australia), restricted the Program to 'Climate Change') and all company statuses were included.[10] I started with 2010, and used `Rvest` to read html entries from each page, scraping each data point. I then repeated this process for each year, and merged the data together. This type of data and full responses on more granular data require an expensive subscription; but the point of the CDP is transparency and visibility, so the public is able to publicly view which companies responded to requests and what their relative score was. Future research could build on this paper through a CDP subscription for more granular data. This data is publicly accessible and anybody can view it, and I have not changed or modified the data, nor am I using it for commercial purposes.

[10] See CDP Data Collection Methods and Description on different categories.

## CDP Data Collection Methods and Description

The way that CDP works is that companies are asked to respond to information requests issued by their stakeholders using the CDP disclosure platform. Organizations that choose to respond to requests without an explicit request by stakeholders can do so as a self-selected company by completing a registration form. The definition of stakeholder is either a shareholder, investor, or a customer. As far as I am aware, the CDP does not publicly disclose which companies are self-selecting and which companies have been request for data. The CDP also does not disclose who requested disclosure. This data would be enormously helpful in validating the causal mechanism in this paper, and answering related questions, like what is the ratio of consumers to investors requesting disclosures? How many companies self-nominate themselves? The CDP may choose to release this information in the future.

The companies that do disclose must provide information on several programs related to the CDP's work; climate change, water security, forests, supply chains, and cities. According to the CDP's scoring methodology, companies are scored according to several items they must disclose within their response, and meet several criteria. To obtain an 'A' scoring, companies must verify Scope 1, 2 and Scope 3 emissions (Is 1.5oC aligned, publicly available, board level oversight and management responsibility, feedback mechanism in place or plans to implement a mechanism within the next two years) is mandatory; in addition, verification needs to cover at least 100% of total Scope 1 and 2 emissions with no significant relevant exclusions (reported in C6.4); and at least 70% of a minimum of one Scope 3 emissions category Companies may have been asked to submit to multiple themes, or a specific theme. Companies are scored on leadership via a questionnaire, and must indicate engagement with suppliers, report a

near-term emissions target that has been validated by the Science Based Targets Initiative or meets the following criteria (Company-wide coverage, coverage of 95% of Scope 1 and 2 base year emissions, a target year that is within 5-10 years of the year the target was set). The scoring itself is sector-based, and therefore bench-marked to the sector.

The CDP scoring evaluates both the extent to which, and the nature of, a company's disclosures on their environmental impacts and progress on climate change. Companies are scored between D-A minus. The decision not use an F grade is not stated, but I imagine this grade would be a deterrent for companies. It is important to note that the scoring methodology used by the CDP is not arbitrary; it is aligned with the Taskforce for Climate-Related Financial Disclosures (TCFD) as well as major environmental standards. Companies are scored on four levels; the composite score is used in this paper as I did not have access to the dissagregated scores. The following descriptions of the scoring methodology are taken from the CDP scoring methods explanation:[11]

**Disclosure (D-/D score)**

Every question in the questionnaires is scored for disclosure. At disclosure level, companies are awarded roughly one point per data point provided. This level features both D and D- score. To score a D over a D- organizations need to have disclosed a more extensive set of information. However, a D-/D score is just the starting point for organizations that want to demonstrate that they have begun their environmental journey.

**Awareness (C-/C score)**

[11] CDP scoring methodology guidance: https://www.cdp.net/en/scores/cdp-scores-explained

A C-/C score indicates awareness-level engagement. Awareness also has C and C- scores, with the differentiator being the level of awareness a company has shown in their response. The awareness score measures the comprehensiveness of a company's evaluation of how environmental issues intersect with its business, and how its operations affect people and ecosystems.

**Management (B-/B score)**

A B score indicates environmental management. Companies that score a B have addressed the environmental impacts of their business and ensure good environmental management. A B-score indicates that a company is showing some evidence of managing its environmental impact but is not undertaking actions that mark it out as a leader in its field.

**Leadership (A score)**

To earn an A score from CDP, organizations must show environmental leadership, disclosing action on climate change, deforestation or water security. They must demonstrate best practice in strategy and action as recognized by frameworks such as the TCFD, Accountability Framework and others. As well as having high scores in all other levels these companies will have undertaken actions such as setting science-based targets, creating a climate transition plan, developing water-related risk assessment strategies, or reporting on deforestation impact for all relevant operations, supply chains and commodities.

**Failure to disclose (F score)**

An F score is given when a requested company fails to disclose through CDP.

The CDP score is based solely on activities and positions disclosed in the CDP response. It therefore does not consider actions or activities not mentioned in the CDP response.

This paper is focused on the climate change program, however future research could expand into Forests and Water Security. For forests, the CDP has used its own disclosure mechanism to track progress on not contributing to deforestation associated with the production of a selection of key agricultural commodities. This measures is now being expanded to avoidance of destruction of all natural ecosystems. This would be a particularly interesting database to expand the findings here, in exploring whether deforestation activities lead to consumer backlash, particularly in regions were logging-activity leads to increases in flood risks. Water security refers to water usage and contamination. Research exploring how coastal populations, or populations with scarce water usage (such as droughts) affects perceptions of company water usage is an interesting avenue of research.

## B2B vs. B2C Distribution

The list of Australian firms in the CDP database is heavily skewed towards Business-to-Business facing firms, with 79.38% of the firms being B2B (Business-to-Business), and about 20.62% are B2C (Business-to-Consumer). This reflects the large energy, agricultural, mining, and construction industries that dominate the Australian market-base. This is not a concern, as the sample is represen-

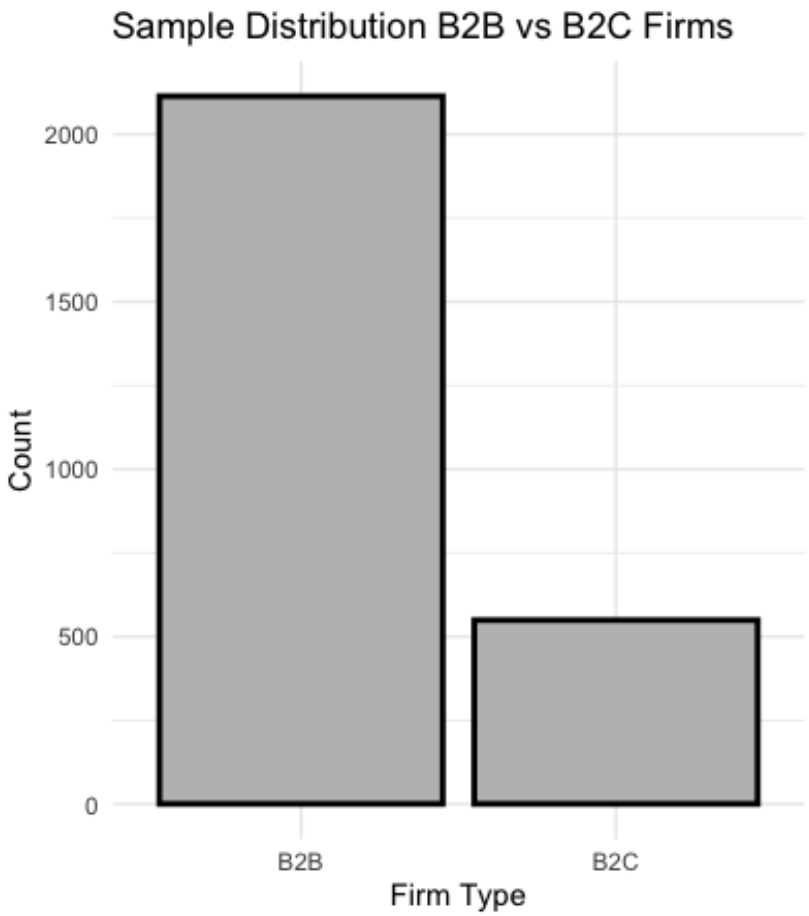



Figure 3: Distribution of B2B and B2C companies in CDP contacted Australian firms

tative of the population of firms in the Australian market-base. Indeed, even in other markets such as the US, B2B companies out-populate B2C companies by a ratio of 1.6:1. It also relates to the manner in which firms are selected to respond to the CDP, as I discuss below.

### Relevance of the CDP to Australia

The Australian government refers to the CDP as Australia's primary voluntary reporting body (CiteDrive 2017). Parliamentary reporting from the Senate Estimates Committee on Carbon Risks cites the Carbon Disclosure Project as the dominant reporting framework for voluntary reporting on emissions. In 2022, the CDP was asked to provide input on the Australian Treasury's consultation on mandatory disclosure of emission in Australia. There has been a rapid growth in reporting to the CDP from Australian companies, the figure below illustrates this growth rate. In 2022, 263 Australian companies disclosed environmental information to CDP, representing a year-on-year growth of 37%. The number of firms approached by the CDP is illustrated in the figure below, showing a steep increase from 2021.

The exponential increase in both companies approached by the CDP and companies disclosing to the CDP in 2022 indicates that number of companies approached explains this increase in disclosures. This is a reassuring sign and suggests that the CDP should expand its requests for submission from companies in Australia.

## Climate-Linked Natural Disasters

In order to establish a stronger link between natural disasters and climate change, to ensure that consumers were reacting to

### Google Trend data by natural disaster

Google Trend data is useful in providing time-series data on fluctuations in public awareness on climate change, especially across more finely-grained periods of disasters. Analyzing whether key search terms related to emissions and disaster

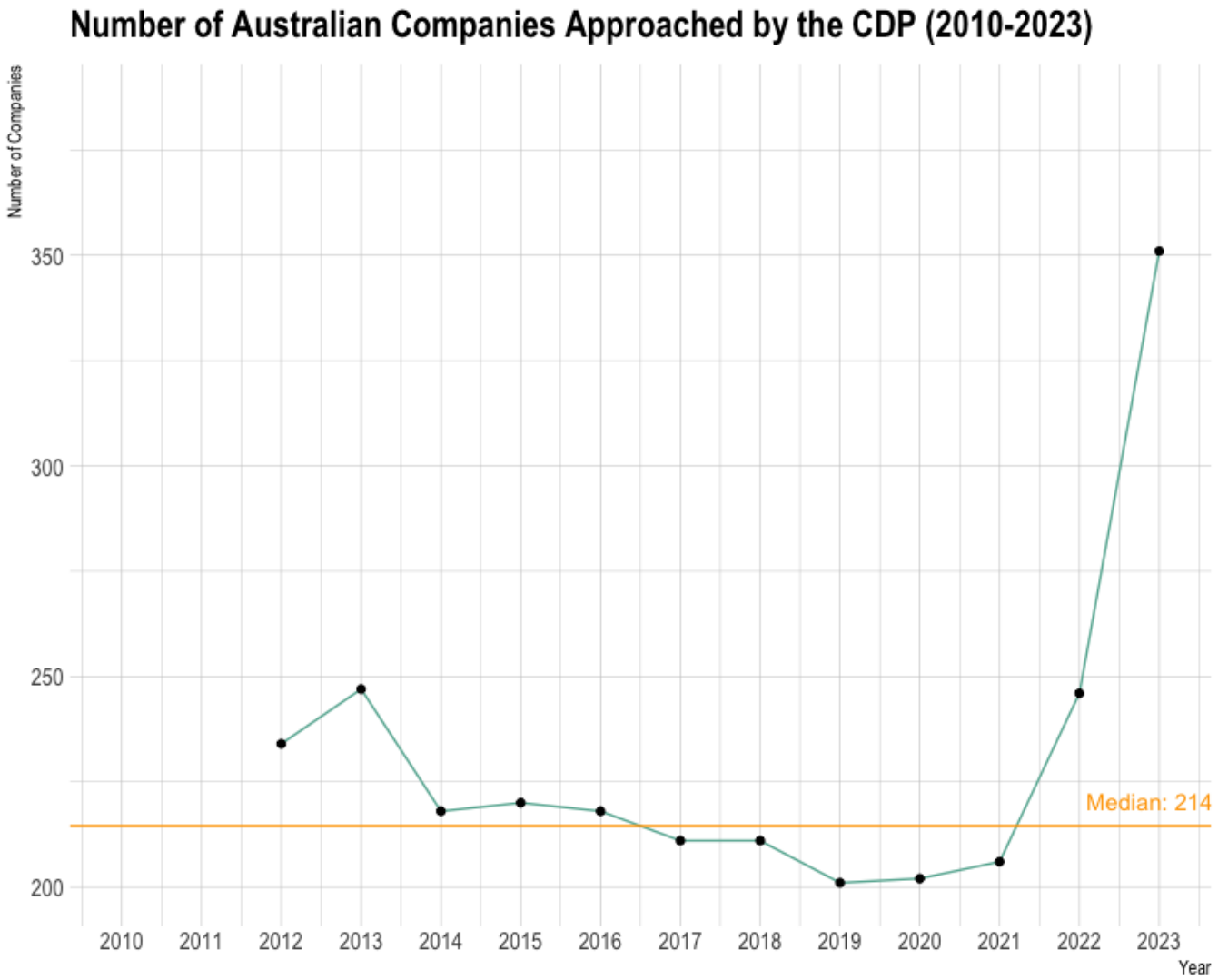



Figure 4: Number of companies approached by the CDP)

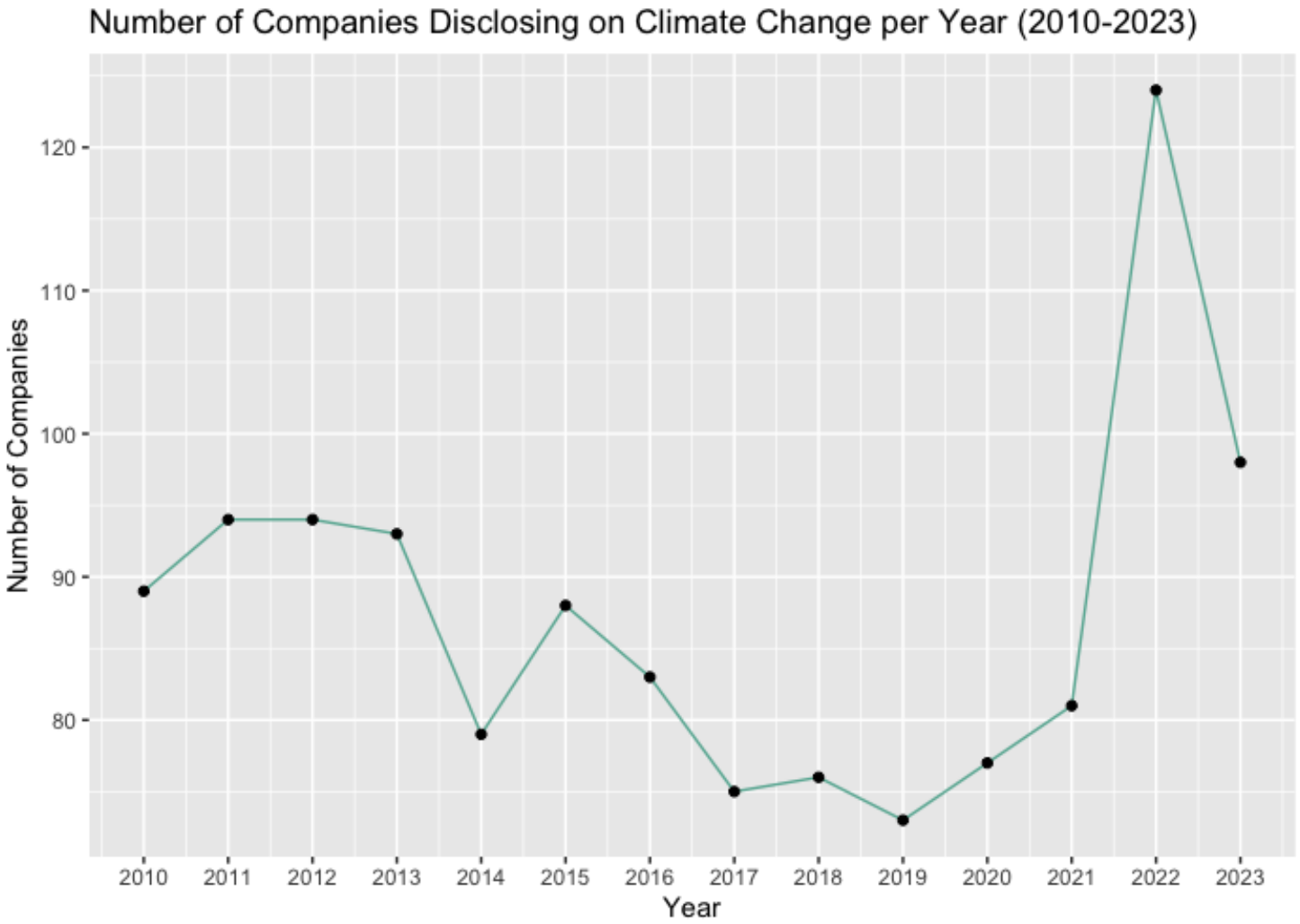



Figure 5: Number of companies approached by the CDP)

move in similar directions can give a broad indication that natural disasters in Australia prompt consumer awareness to emissions. As per Google Trends, this software, "normalizes search data to make comparisons between terms easier. Search results are normalized to the time and location of a query by the following process: Each data point is divided by the total searches of the geography and time range it represents to compare relative popularity".

The use of online queries for environmental analysis is a growing field. While there is a paucity of studies that specifically focus on how consumers perceive greenhouse gas emissions online or the climate-related behaviour of companies, an ample number of studies have examined public awareness of climate change. One such significant study was conducted by Kirilenko, Molodtsova, and Stepchenkova (2015), demonstrating a link between unusual temperature patterns and the frequency of tweeting. Beyond the realm of climate change, the topic of energy consumption has been investigated through the lens of Twitter. Holmberg and Hellsten (2015) analyzed Twitter dialogues pertinent to energy, aiming to decipher patterns in energy consumption behaviors and the impact of social media discussions on these trends. Building on this, Kryvasheyeu et al. (2016) conducted research showing a direct connection between the level of Twitter activity and environmental disasters in the United States, indicating that environmental threats can be observed through increased Twitter discussions. In a similar vein, Sisco, Bosetti, and Weber (2017) used Twitter data to examine how short-term public awareness of climate change responds to extreme environmental events. Their findings revealed a correlation between public attention and both the financial impacts of climate events and the extent of their deviation from the norm. This would suggest that social scientist may not need to focus on confirmation attribution when it comes to scientifically

linking natural disasters to climate change, rather the evidence would suggest the connotation between natural disasters as causal evidence of climate change is so strong in public discourse that any natural disaster will prompt or prime public awareness of climate change. The finding that this effect is acute for the financial aspects of climate change and the heavy costs associated with it is promising, and reinforces the methodology utilised in this paper in using insurance data to measure high-impact climate related disasters.

I used google search results over the course of a 10 year period for the key term 'emissions' across sub-regions and cities in Australia. This follows the work of Zhang, Abbas, and Iqbal (2022) who used terms related to 'GhG emissions and 'renewable energy' to test whether the public's receptiveness to environmental policies fluctuates in response to extreme environmental circumstances. I dropped 'GhG' because individuals are more likely to use the term emissions than adding GhG. As an additional validification test I examine related queries and related topics across the continent and sub-regions. Related queries and related topics represent terms that were searched for alongside the term of interest; the top terms had the highest growth in volume over the time period of interest. The reason this is significant is that it may indicate a spurious or non-related search for emissions in a region that is not linked to climate change emissions. The top related topics are reassuringly linked tightly to emissions, including: 'greenhouse gas emissions', 'cut', 'paris agreement', 'emission', 'important', 'environment', 'carbon pricing'. The top related queries include: 'zero net emissions', 'paris agreement', 'scandal', 'climate change definition', 'zero emissions'. These are topics queries that reinforce the contextual search for emissions.

The figure below shows Google search trend results between disaster search and emission searches, the co-variation of the lines is a promising sign that individuals are primed to think about potential contributing factors to natural disasters, and related disasters. Observationally, the spikes in both emissions and disaster key search terms co-vary with climate-linked disasters; for instance the huge spikes in 2016 makes sense in the context of the so-called 'angry summer' where over 205 records were broken across Australia with successive heat-waves. New South Wales mean temperature in summer was on record with the highest temperatures recorded at 2.57 degrees celcius above average. 2016 also involved high levels of flooding, fires, and drought.

The search above is promising but the problem is we cannot track whether there is a relationship between regions hit by disasters and this trend. In some ways this is to be expected; large-scale natural disasters in Australia receive wide-coverage, with families split across the continent, it would make sense for search results on emissions to be high in Western Australia even if the disaster had hit the East Coast of Australia. However, to further probe the relationship between searches for emissions and damage from climate-linked natural disasters, I geo-code damage from natural disasters in QGIS using a shape file, and overlay this with google-trend data of the term 'emissions' by sub-region and city for the 2019/2020 black summer period where bush-fires impacted almost every state in Australia over the course of several months, increasing in severity from November 2019 until February 2020. Climate-linked natural disasters are not just damaging in the first instance, bush-fires have devastating and long-lasting effects on the environment and individuals livelihoods. To measure damage from natural disasters, I focus on bush-fires, and use the National Indicative Aggregated Fire Extent Database. This data was developed by the

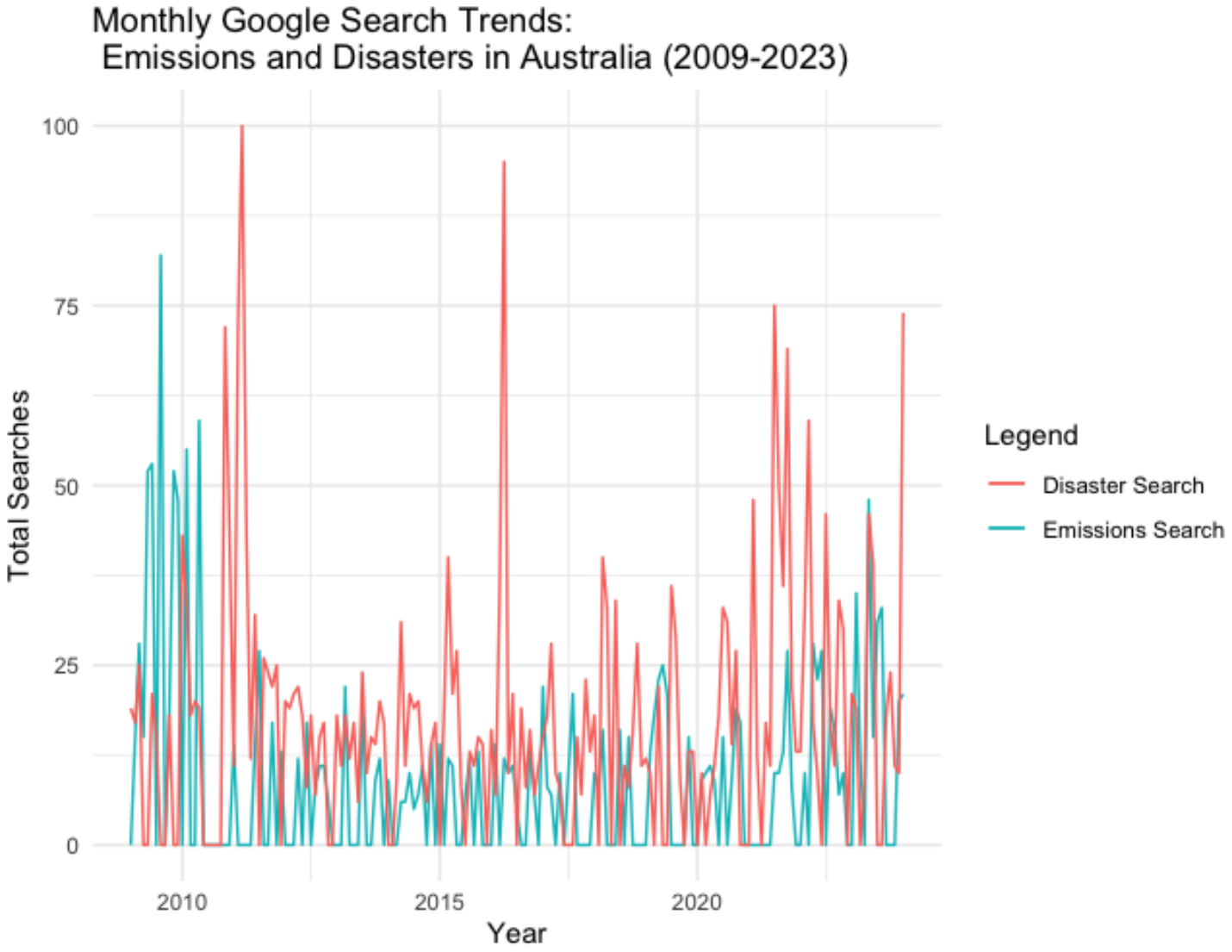



Figure 6: Google Trends data results 2010-2023 on Emissions Searches and Disaster Searches)

Department of Climate Change, Energy, the Environment and Water and was developed to quantify the impacts of the 2019/2020 fires on wildlife, plants, and ecological communities. The shapefile indicates burnt areas across Australia for the 2019/20 bushfire season, with gradient colour that indicates severity of the damage. The data covers July 1st 2019 until 22 June 2020. The dataset adopts the national Emergency Management Spatial Information Network Australia (EMSINA) data service, the official measurement of fire damage, with additional data added.[12]

The figure below illustrates the Map of Australian city search terms for 'emissions' and the severity of the burnt land. The green points indicate high-search term areas for 'emissions' between November 2019 and February 2020. The sparsity in Darwin is related to internet access in the Northern Territory, however, the two points across cities in the NT are in high-fire exposed regions. Similarly, the East Coast regions that were hit show a greater clustering of search terms.

[12] An additional note here, when comparing popular search trends during 2019/2020 for both 'fires' and 'climate', a top 'breakout phrase' used was "is climate change causing fires in Australia". This is strong evidence that this period of time in particular involved high priming of individuals between natural disasters and climate change.

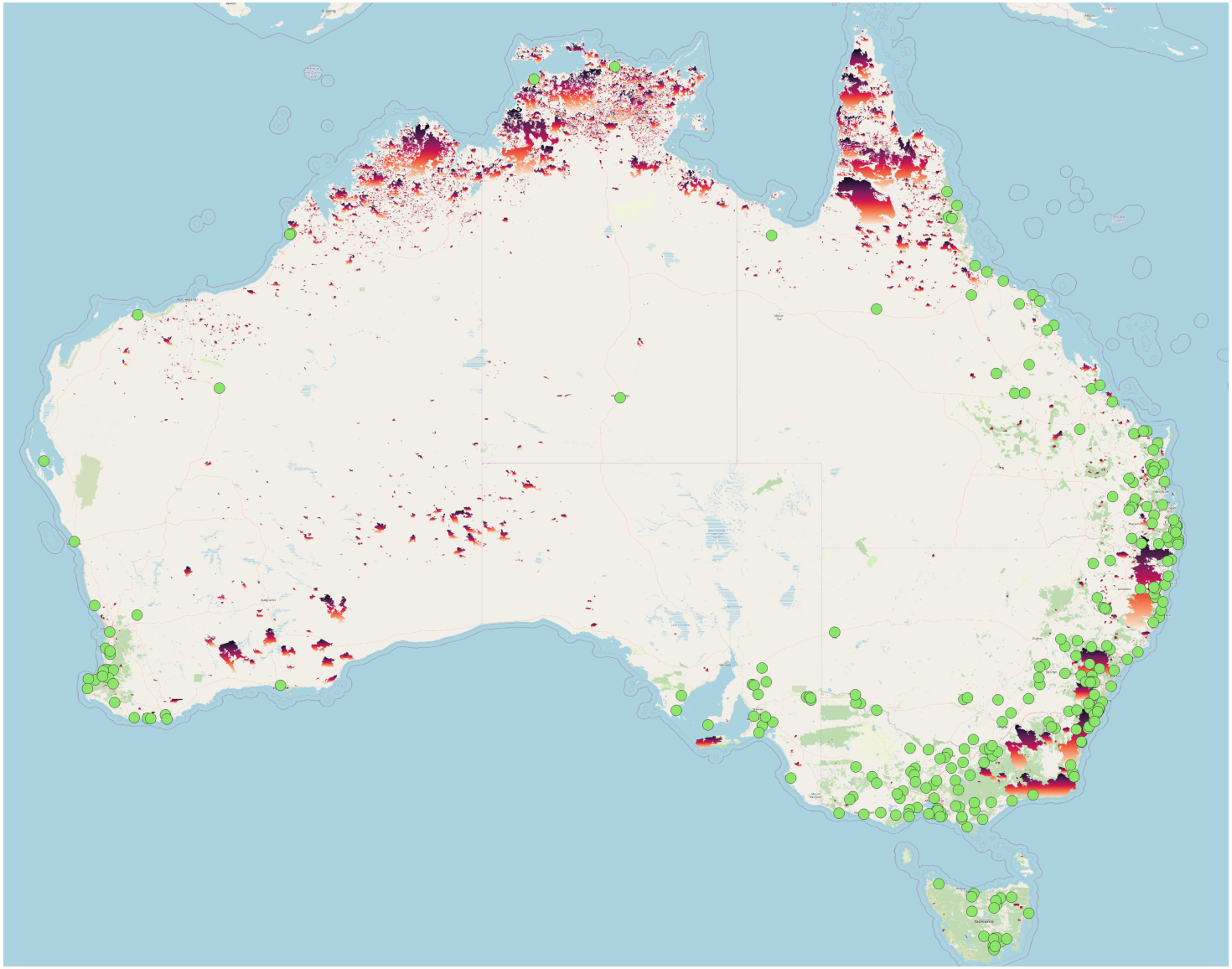


Figure 7: Geo-Coded data from QGIS Tracking burnt land across Australia from the 2019-2020 bush fires and google search trends for 'Emissions'. For reference, google trends classifies a keyword or phrase as 'breakout' "if a topic or query that has recently grown more than 5,000% in search volume".

## GDELT Climate Narratives Database

A huge problem in the scientific literature is confirmation attribution for extreme climate events. That is, how can we know that a natural disaster or extreme weather event was caused by climate change instead of just natural weather patterns? Particularly in Australia, bushfires are endemic, made worse by highly flammable dry-lands and eucalyptus trees. Heat waves are part of summer, and it is often difficult for the public to discern between normal weather patterns and abnormal weather patterns. In some ways, the job is much easier for the social scientist, because it is not so important for us to prove that an extreme weather event was the result of changes in temperature of theta-e, but rather it is more important to prove that people *think* that an extreme weather event was caused by climate change. The perception that this event is linked is the more important element to my argument; if my argument is that companies are conscious of their visible environmental-related behaviour during 'climate shocks' then it must be the case that consumers are truly being primed during natural disasters. I cross-referenced the disaster years I used as high and low disaster years for events by using Google news aggregates of articles; how many articles made references to climate change in the context of a given disaster, and the number of publications covering this topic. One limitation here that is deserving of criticism is that, as I argued above, the link between natural disasters and climate change has grown so strong in the last decade of my sample that many articles make at least passing references to climate change in the context of an extreme weather event. I therefore turn to the GDELT (Global Datasets on Events Location and Tone) climate change narratives database. The GDELT dataset covers data on of 6,092,147 linguistically annotated 5-word snippets of climate change mentions, drawn from our 101-billion-word Web Part Of Speech dataset. Recent work using GDelt data with links to natural disasters

includes Djourelova et al. (2023),

One limitation here is that the only data for GDELT that is available is 2015-2020, here I just consider whether media coverage of climate change and language around emissions peaks during natural disasters.

## Robustness Tests

I add additional fixed-effects to my models with year, industry, and company fixed-effects. This multi-level fixed-effects approach significantly enhances the robustness of the analysis, enabling me to isolate the effect of being a B2C firm on the likelihood of carbon disclosure, thereby providing more accurate insights into the interplay between market orientation and environmental transparency in the corporate sector.

Table 6: Probit Models: Effect of B2C Facing Firms on Emissions Disclosure with Three-Way Fixed-Effects

| | *Dependent variable:* | |
|---|---|---|
| | Disclosure Dummy | |
| | *OLS* | *probit* |
| | (1) | (2) |
| Business to Consumer | $-0.070^{***}$ (0.022) | $-1.377 \times 10^{15***}$ ($6.855 \times 10^{7}$) |
| ESG Score | | $2.631 \times 10^{13***}$ ($2.346 \times 10^{5}$) |
| Direct Emissions | | $4.878 \times 10^{13***}$ ($4.868 \times 10^{5}$) |
| Company Market Cap (USD) | | $6.728 \times 10^{4***}$ ($1.415 \times 10^{-3}$) |
| Company Size: Medium | | $-6.487 \times 10^{14***}$ ($1.220 \times 10^{7}$) |
| Company Size: Large | | $-1.287 \times 10^{15***}$ ($1.494 \times 10^{7}$) |
| Constant | $0.393^{***}$ (0.011) | $-9.790 \times 10^{15***}$ ($8.994 \times 10^{7}$) |
| FE: Firm, Industry, Year | *No* | *Yes* |
| Observations | 2,663 | 1,377 |
| $R^2$ | 0.004 | |
| Adjusted $R^2$ | 0.003 | |
| Log Likelihood | | -6,271.596 |
| Akaike Inf. Crit. | | 12,953.190 |
| Residual Std. Error | 0.483 (df = 2661) | |
| F Statistic | $9.657^{***}$ (df = 1; 2661) | |
| *Note:* | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

# Discussion and Conclusions

## 1 The Reputation Variable in IR

In Nine Lives, Plato tells the story of an Athenian demagogue and leader, Alcibiades, who was known to love animals, and he himself purchased a dog for the steep price of 70 minas (Plutarch et al. 1960). The dog was known to be his "principal ornament", but Alcibiades cut off the tail of this dog. Plutarch suggests it was for the sole reason that it would be so shocking to Greek observers that anything he did in the future would pale in comparison, and thus the example of his extreme behavior would dampen the social effect of any future bad behavior. His reputation would be forever 'bench-marked' to that time he cut off his dog's tail, making all his future deeds pale in comparison. Reputation has a similar function in political life; it is not only a concept that largely depends on the historical reputation of the bearer, but also the behavior of their peers. Reputations are not fixed but rather bench-marked to a reference point and are, as I argue in this dissertation, decided largely by the group's moving average. It's a theme that echoes throughout the history of International Relations scholarship, where reputation takes different forms, and possesses different power, depending on the value the audience places on that particular type of reputation at that point in time, as well as the extent to which the owner of that reputation's survival depends on maintaining a particular reputational type among key audiences.

This dissertation puts forward a competing view of reputation as a group-based concept. Whereas classical theories of reputational costs tend to treat reputations as a uni-dimensional idea that emphasizes a singular trait across all contexts - usually *resolve* - I suggest that certain reputational traits are valued depending on the situation, and the priorities of the audience. This suggests that actors can have multidimensional reputations, and indeed, some of these traits can be competing and conflicting if the actor is managing multiple audiences. The timeworn debate that adjudicates between reputational models that are situational, which is to say a property of the environment that actors asses in the same way, and reputational models that are dispositional, where actors draw different conclusions, does not capture how reputations do not sit in vacuums. I argue that both the audience themselves, and the bearer of a reputation, are influenced by the position and judgments of their group. The reputation of the bearer depends on the actions of their peers, and the judgment of the audience depends on the judgment of their peers.

Despite the extensive work of Robert Jervis on the concept of reputation, he was clear that his treatment of reputation was necessarily stripped down, and that the problem of the individual, versus the group, and *conflicting* reputational types among actors posed significant questions to the discipline. In a footnote of his own dissertation, Jervis remarks:

> "An actor's reputation for issuing accurate signals is only part of his general reputation, and image. An actor may be thought to be unwilling to run high risks and also believed to almost always act with his signals. Or he may be thought to, be both a bluffer and a high risk taker. Or, to take another characteristic, he may

> be thought to issue accurate signals yet be aggressive. And in any given situation a state's signaling reputation is not the only part of its image which determined others' predictions of its behavior. As we previously noted, decision-makers may take bluffing as an index to other characteristics, and other characteristics may be taken as an index of whether the state will bluff. Thus the state's reputation for issuing accurate signals may not be totally independent of the rest of its image. This question requires further investigation but will not be considered here. For the sake of brevity we will sometimes use the word reputation to mean only reputation for issuing accurate signals"(R. L. Jervis 1968, p. 71).

The idea of many reputations, and how they interact, and when they become salient, can be found in an original IR text of Hobbes.

In chapter 10 of Leviathan, 'Of Power, Worth, Dignity, Honour, and Worthiness', Hobbes writes:

> "Reputation of power, is Power; because it draweth with it the adherence of those that need protection. So is Reputation of love of a man's Country, (called Popularity), for the same Reason. Also, what quality soever maketh a man beloved, or feared of many; or the reputation of such quality, is Power; because it is a means to have the assistance, and service of many. Good success is Power; because it maketh reputation of Wisdome, or food fortune; which makes men either fear him, or rely on him. Reputation of Prudence in the conduct of Peace or War, is Power; because to prudent men, we commit the government of our selves, more willingly than to others" (Hobbes n.d.).

Hobbes is making the point that a reputation can imbue power in an actor and, rather than stopping at the idea that reputation's function is limited to a credible commitment by showing resolve and deterring aggression or displaying honor, Hobbes goes on to speak of different types of reputations and consequently, the different types of power this generates for actors that build these reputations, and for audiences that value them.[1] Given the significance in which IR scholarship has divided itself up into theoretical, methodological, and disciplinary camps (Lake 2013), the predominance of reputation in IR as a catch-all variable is surprising. On the one hand, reputation is used as a vague panacea for explaining a wide variety of phenomenon; is the celebrated hero in all sectors of the field, featuring prominently in everything from why states go to war, why they cooperate, and why they might comply with international laws. On the other hand, as this dissertation argues, despite how ubiquitous reputation has been in the field of IR, reputation has been narrowly applied within the rigid structure of states with a focus on resolve. This dissertation has steered away from the classical application of resolve as a type of reputation, the most over-saturated aspect of reputation, and states as the major actors.

The growing power of firms as political actors in IR provides an opportunity for international relations to expand its conceptualization of reputation by learning from the fields of micro-economics and management in how they have dealt with firm reputation. The first part of this dissertation introduces an overarching theoretical framework to understand how the different disciplines developed ideas around reputation and why they evolved differently. I then offer an overarching theoretical framework to understand reputation as being group-based, rather than just historically-dependent.

[1] Of course, Hobbes regarded reputation as one of the three causes of fighting in IR - i.e. the glory for reputation.

The disputed definitions of reputation encapsulate the debate on the necessity of history for reputation. For example, Dafoe, Renshon, and Huth (2014) defines reputation as "any belief about a trait or behavioral tendency of an actor based on past actions", whereas Renshon, Yarhi-Milo, and Kertzer (2023) use a more socially-conscious reputation with "socially shared beliefs about an actor's behavioral tendencies", and Yarhi-Milo (2018) opts for a vaguer definition that divorces reputation from the necessity of historical behavior, here reputation is, "the belief that others hold about a particular actor". The variation in whether reputation requires observation of past actions is clear in these three definitions. Little work has interrogated the history-dependent or 'vertical' comparison of reputation – i.e. the history of an actor or subject's reputation where you can compare along different points in time to assess the relative effect of an action on the subject's reputation. Even research that seeks to prove that history does not matter for how observer's make judgments conclude that this proves reputation doesn't matter, rather than exploring whether there is another way for observer's to assess reputation.

Less common in IR is what could be called the 'horizontal' comparison of reputation, or a group-based concept of reputation – that is, the actor's reputation is not assessed in comparison to their history, but in relation to their peers. The other point of contention highlighted by these definitional differences is whether the beliefs others hold need to be socially-shared by a group - or just a single belief held. My dissertation finds that none of these definitions satisfy reputation, but a broader definition, such as Yarhi-Milo (ibid.) applies best. I offer a competing definition, that reputation is "beliefs that a group shares about an actor's behavior in a setting". This dissertation suggests that the

context and domain of behavior shape which type of reputation matters, and while shared beliefs among the observers of a reputation is a necessary criteria of reputation, these beliefs might differ between groups.

Alongside the theoretical issues that underpin reputation, The empirical landscape of reputation is unevenly traversed. Reputation is often used in qualitative work since it is a difficult concept to quantify; and indeed where it has been studied empirically, measurements of reputation must invoke it at a distance, or presume it exists tautologically when we attempt to explain or understand different outcomes, as is the case in the audience cost literature, or studies on compliance with international law. The theoretical dimension of how reputations are made, maintained, or broken – and the thresholds under which this occurs is still unclear, as is the conditions under which some reputations matter to some actors more than others.

Part of the source of this problem is that reputation is a concept with great purchase outside of the field of International Relations. John Ruggie once wrote that "methodology can become tacit ontology", and when IR scholars have one understanding of what reputation is, what type is most important, and how it is observed, the methods we use to understand it can subtly and often unconsciously shape our understanding of what it is. In this dissertation I sought to address reputation by connecting it with other strands of work on reputation in the fields of political psychology (R. Jervis 1989), labor economics (Tirole 1996), behavioral economics (Kahneman, Sibony, and Sunstein 2021; Kahneman, Knetsch, Thaler, et al. 1986) and behavioral science (Kuran and Sunstein 1998), finance (Roberts and Dowling 2002), environmental science (Milinski, Semmann, and Krambeck 2002), and organizational psychology (Fombrun and

Shanley 1990). While these fields might seem disparate or outside the expertise of IR; as long as research is connected to reputation, this research offers enormous insights. Robert Jervis, who has pioneered the field of political psychology for IR theorists, made the point that many strands of literature are often kept separate, when they really depend on each other (R. Jervis 2017, p. 11). No more is this true than the study of reputation which, at its heart, like Hobbes idea on the state of nature, began as a microcosm of individual relationships that was scaled up to discuss state behavior.

In bridging these fields together, the large gaps in the literature on reputation sit in five camps. They are, 1) reputational types, 2) the outcomes that reputation produces in different contexts, 3) the necessity of crisis in the reputational literature, 4) can non-state actors develop strong reputations, 5) and the existence of multiple, competing audiences with different expectations over reputations. These gaps formed the basis of the three puzzles I address in this dissertations; A) How does the group-effects of reputation affect firm behavior B) What effect do different types of crisis across domains of IR have on generating reputational costs and how actors respond? and C) How do agents strategically manage multiple audiences and conflicting reputations among those audiences? I departed swiftly from the IR preoccupation with resolve in the context of war and conflict signaling, and pivoted away from the field of international law and state-compliance. Instead, this dissertation focuses on the role of firms, and empirically on two cooperation problems in international political economy; geopolitics and business, and climate change and climate governance.

Using ideas fermented in different silos of academic disciplines, I argued that the same social dynamics that behavioral scientists and economists observe oc-

curring when individuals form groups, or make decisions in group settings, also conditions the way that non-state actors - firms - make decisions, and thus can be used to explain why firm's might choose to cooperate, self-regulate, or change their behavior. I argue that IR's pathological obsession with states as the major actors, war as the only domain, resolve as the dominant reputation worth fighting for, and audiences as a singular monolith imposing one large cost at one time of crisis, has confined the study of reputation in IR to a static, singular phenomenon that defines the limits of questions over the extent to which an actor's concern for their reputation can affect outcomes in world politics. Since the unspoken but collective consensus that, under conditions of anarchy with little enforcement mechanisms available that are independently wielded from state power, reputation is the only true enforcer in the system of states, little progress has been made to uncover the many facets of reputation as a variable in IR. This dissertation offers an overarching framework of a group-based concept of reputation to explain how firm concern for reputation can affect behavior. In the two empirical papers, paper 2 and paper 3, I tease out some of the ideas developed in paper 1 in different domains, including the role of dueling audiences, incremental costs, and different reputational types. Table 1 summarizes the different approaches of each paper in the theme, the two audiences, the actor type, the reputational type, and the reputational cost involved in each case.

| Overview of Reputation across Two Papers | | |
|---|---|---|
| Domain | Geopolitics & Sanctions | Climate Change |
| Duel Audiences | Public vs Public | Public vs Shareholders |
| Actor Type | Firms | Firms |
| Incremental Cost | Russian authority sanctions, Russian consumer backlash | Shareholders penalty, Public backlash |
| Reputation Type | Human Rights | Environmental |

Table 1: Incremental Sanctioning

Table 2: Comparison of Results: Dependent Variables and Mechanisms

| **Dependent variable** | **Human Rights** | **Climate** |
|---|---|---|
| **Variables** | | |
| Reputational Costs | ✓ | ✓ |
| Disclosure | ✓ | ✓ |
| Commitment (signal) | ⋆ | ⋆ |
| Incremental Costs | ✓ | ✓ |
| Duel Audience Costs | ✓ | ✓ |

The main empirical findings on reputational costs and the mechanisms through which it affects behavior are summarized in Table 2. On the whole, Table 2 offers relatively consistent evidence that concern for reputation can induce agents to disclose or to use signaling devices to appease audiences. As evident in Table 2, dueling audiences are present in paper two and three - and the bearer of a reputation(s) must deal with suffering reputational costs among one audience in order to appease another. Table 2 also shows that this results in the use of covert strategies to appease both audiences; for geopolitics, war and human rights in paper 2, this involves hedging bets and buying time by issuing statements to appease international and western consumers, while maintaining operations in Russia and avoiding politically charged language to satisfy Russian audiences. In the context of climate change, this means firms may choose to disclose their environmental practices to appeal to consumers and the public, but only if they are confident of receiving a high score. Alternatively, they may opt to disclose while requesting that their grading remains confidential. Table 2 also suggests support for the role of incremental reputational costs at motivating actor's to change behavior; rather than a single large sanction, cumulative sanctions or fear of cumulative sanctions, can push actor's to change their behavior. Table 2 demonstrates that both papers 2 and 3 find support for the role of crisis; crisis is a necessary ingredient that amplifies the effects of reputational costs by increasing awareness of an actors behavior and reputation; for paper 2 this is war and social media attention, and for paper 3 this is natural disasters and extreme weather events.

Cumulative sanctions and incremental sanctions help explain the cascade effect; where an early-mover can trigger large-scale judgments and statecraft movements in which actors begin to believe something, or do something simply

because others appear to (Sunstein 2019, pp. 18–19). A key contribution of this dissertation was to shed light on the mechanism behind reputation - where we can identify that reputational costs may be motivating an actor's behavior, and to help explain why reputation often manifests in the way that it does - starting with the actions of one actor, involving a huge wave of change, but then the movement is perhaps ultimately short-lived. The findings in paper two, and to some extent paper three, show that reputational cascades can increase the pace of an actor's behavior or decisions within a group, and can help explain the timing of behavior. While firms withdraw quickly following the actions of an early mover, this pace ultimately slows over time. Paper three suggests that for some areas of IPE, such as climate change, the nature of the crisis in this area creates priming events that are more likely than in other domains. That is to say, natural disasters are increasingly more common than say, war. This means that cascades are more likely to happen more often, and even as consumer pressure eases as the salience of the event passes and consumers perhaps lose interest in the environmental behavior of firms, they are likely to be primed again in the future. This is of course a geographically determined phenomenon. My case study focuses on Australia, and the trend may only hold in countries or regions that face similar exposure to the effects of climate-change induced extreme weather events.

A future line of research could build on this to help explain why some firms (or other types of actors), are more likely to be susceptible to a reputational cascade while other actors are not. Clearly the stakes of the actor (how much they stand to lose) matters. Second, the power and resources of the actor (how much resources they have to withstand the backlash),[2] and how visible they

[2]It does not always hold, as is the case with states as well, that more powerful actors are less sensitive to reputational costs. It is the relationship between the actor and a particular audience(s) that is critical.

are to the audience (business to consumer or business to business facing firms) matters. Future research might uncover how important or critical these characteristics of firms are in their susceptibility to consumer pressure.

Table 2 demonstrates that two papers find support for the role of crisis; crisis is a necessary ingredient that amplifies the effects of reputational costs by increasing awareness of an actors behavior; for paper two this is war, and for paper three this is natural disasters and extreme weather events. For paper two and three, it is clear that media attention to these crises is also critical in fueling the crisis, raising audience awareness, and reducing informational asymmetries

## 2 Reputation for Non-State Actors in International Political Economy

The focus on reputation in conflict-dominant areas of IR has its origins in the how reputation entered the discourse of IR; emerging out of Cold War scholarship on deterrence, and the Schelling hypothesis that the West would need to demonstrate resolve in peripheral fights with the Soviet Union. Without this credible commitment and demonstration of 'brinkmanship', Soviet aggression would not be contained (Weisiger and Yarhi-Milo 2015; Schelling 2001; Schelling 1966). In the context of war and conflict, reputations for resolve - "the extent to which a state will risk war to achieve its objectives" (Mercer 2010) - are critical for a state's ability to signal its willingness to risk war, and thus their ability to deter aggression. In this context, the only actor that matters is the state, and the only way to acquire a reputation for resolve is through past behavior. Outside of this context, however, there is reason to assume that reputations among non-state actors, in domains *other* than conflict signaling and war, could require

new assumptions about how these reputations are formed and maintained.

The importance of reputation to an actor depends in part on the type of reputation they're concerned with cultivating, whether or not it is in conflict with another salient reputational-type, and the nature of the system in which they operate. Thus, the logic of reputation must be adapted when we're examining different types of reputations and different types of actors. As Robert Jervis argued, "The importance to an actor of his reputation depends partly on the system within which he is operating. Being caught lying generally has lower costs in international relations than in the domestic politics of most states. In the latter outright lies of any sort are not expected and their occurrence therefore immediately casts grave doubts on all of the actor's other signals" (R. L. Jervis 1968, p. 72). Here the reputation for honesty, and credibility, is the priority. As the setting changes, so does the importance of the variety of reputation. While Jervis distinguishes between just two settings - the domestic and the international - the context changes depending on the nature of the issue or cooperation problem.

In labor-management relations for instance, a reputation for honesty is often of the utmost importance. While this statement could not be applied to the entire international arena where trust is low and deception is expected from states, it's still an important reputation between allies, and for other actors where membership to the organization, or perhaps stakes in the organization, are voluntary, trust is a very important reputational type. Reputations can also be less clear-cut than the literature on resolve has suggested, rather than reputations for resolve, or compliance, actor's can develop more moralistic, amorphous reputations among audiences - as 'good' types and 'bad' types. Of course these

types might bundle up reputations - as is the case in paper two with corporate reputations for human rights abiding and honesty, or paper three, with environmentalism and compliance. By introducing the dimension of multiple audiences, this dissertation has also suggested that actor's may cultivate conflicting reputations to appease audiences, utilizing specific strategies to achieve this, as I discuss further below.

.

## 3 The Necessity of Crisis

A critical implication of the findings in these chapters is that crisis is an essential ingredient for the role of reputation by increasing i) attention of the audience to a particular type of reputation that, absent the crisis, they may not care about, and ii) increasing the costs for the actors' decision-making. In summary, crisis works to increase the relevancy and stakes of a reputation. This doesn't necessarily challenge standard understandings of the role of reputation in IR, in that the literature on the reputational type of *resolve* requires crisis, but much of that reasoning has to do with the contextual factors that necessitate applications of resolve. It has not yet been suggested that *all* reputational types require some type of crisis that specifically calls into question an actor's reputation in that domain. Crisis helps explain why some reputations matter in some situations, whereas others do not - or other reputations become less relevant. My findings suggest this is the case; not that reputation doesn't consistently matter for actors, but that audiences are only primed to care about certain types of reputations, and those reputational types are 'activated' by that awareness during times of crisis. The crisis itself is merely a way of increasing transparency and lowering informational asymmetries between audiences, it also explains why reputation is a temporal phenomenon - it's effects wain over time as attention

and salience decrease.

The findings advanced in this dissertation suggest crisis matters not just for increasing the stakes of reputational costs by allowing an opportunity for actor's to tie their hands, but more so by priming audiences, and increasing saliency, transparency, and creating conditions conducive to cascade effects. An interesting avenue for future research would be to explore whether the nature of the crisis matters for some reputational types more than others, and even more importantly, how we can measure a crisis. Do larger crisis have greater effects on reputation?

## 4 Domains of Reputation and its effectiveness

This dissertation presents a group-based theory of reputation and empirically explores these specific mechanisms in two areas of international political economy. This explicit focus on IPE is a deliberate move away from the focus on resolve in IR. By focusing on IPE, this dissertation is able to interrogate whether reputation can induce behavioral change in IR among firms in lower-stake domains. The findings suggest reputation has this power, albeit with caveats. Like states, and the succession of literature that has found autocratic states to be more insulated from reputational costs than democratic states (Potter and Baum 2014; Slantchev 2012) due to selectorate theory (De Mesquita et al. 2004; Weeks 2012), firms that rely less on certain types of audiences are also more insulated from reputational costs. For instance, paper 2 suggests that Russian-allied firms are less sensitive to reputational costs, and paper 3 shows that business-to-business facing firms face less costs relative to business-to-consumer facing firms. These findings on which characteristics of firms or situations affect the power of reputation in motivating behavior contributes to the literature on reputation in IR.

Below I discuss two aspects of the findings in this dissertation for factors affecting the relevancy and potency of concern for reputation in motivating behavior: the type of actor involved, and the nature of the costs (incremental costs).

## Reputation and incremental costs

One of the other well-worn foundational areas of theories of reputation in IR is the punishment or penalty imposed on the actor for violating their reputational type. Actor A commits action X which violates their reputational type, and an audience imposes a singular 'cost' on the actor. In most theories of resolve this involves one large sanctioning effort - perhaps the domestic audience votes out the leader who failed to show resolve after signaling their commitment to fight. Alternatively, perhaps the international audience refuses to cooperate with the actor that failed to comply with an international agreement or treaty. The sanction is all or nothing - if the leader is voted out of office then there was a cost to them violating their reputational type, but if the leader remains in office, there is no penalty. Or perhaps in the case of the company that suffered reputational costs from remaining in Russia, or failing to signal action on climate change, did they suffer some large decreases in profit that threatened their continual survival? This example might be classified as a 'heroic' sanctioning effect that illustrates the effects of reputational costs on the actor.

These examples help illustrate the effect of reputation in conditioning actors' behavior; either by guiding their action, or explaining why an actor might be sanctioned. However these examples also obfuscate instances where actors are guided by smaller and less absolute costs that accumulate over time. Costs

are incurred by political leaders, not just through elections held at various time points, but also through shifts in public support between elections. Similarly, costs are felt by companies, not just through cuts in profit, but through slander or misinformation that may affect their brand image - foretelling some future effect on profit even if it's not realized immediately. What I mean by this is that there are 'incremental sanctioning' effects (Coleman 1994), whereby an actor known to have a particular reputational type, commits action X that violates this type, and certain members of the audience impose smaller retaliatory costs on the actor, however these costs are additive and contagious and ultimately have a larger reputational cost on the actor. This idea of 'incremental sanctioning' in the sociological literature demonstrates how norms are enforced, not through one large sanctioning effort, but by smaller efforts over time (ibid.).

By moving away from IR's emphasis on reputations for resolve in settings of war, with the major actors as states, I incorporate ideas endemic in sociological and behavioral literature; that reputation is a group-based concept that develops in reaction to how audience(s) observe relative behavior to both history and peers. Reputations are group-based and they are behavioral; rather than fixed, they change and adapt to audience expectations, and they are far from one dimensional. The costs are not absolute as IR has suggested. For instance, the leader is voted in or out of office, the state becomes known as an unreliable actor, or a liar, or states stop attempting to cooperate with the actor after one demonstration of non-compliance, defection, or cheating. In reality, the costs are usually incremental. As I argue in these papers, the size of the costs usual results from the network of individuals, or the size of the audience, that collectively react to the actor's violation of their reputational type. Just as the network of peers helps determine the reputation of the actor, so too does

the network of the audience determine the costs imposed. I referred to this at the beginning of this dissertation as 'incremental sanctioning'. The table below attempts to illustrate this effect across the three papers:

| Incremental Sanctioning Comparison | | |
|---|---|---|
| IPE Area | Geopolitics & Sanctions | Climate Change |
| Audience A | Consumer backlash or boycott | Consumer backlash |
| Audience B | Russian authority sanctions, Russian consumer backlash | Shareholders penalty |

Table 3: Incremental Sanctioning

In the above table I split the incremental sanctioning between the dueling audiences; each audience may impose a penalty depending on the type of reputation they believe has been violated. The incremental sanctioning in paper two involves consumer backlash. Heroic punishment here might be extreme loss of profit, where it's unlikely that the consumer backlash or boycotts would deliver this at that scale, the brand's reputation is affected by the incremental penalties imposed by consumers online and media coverage.

## Reputational change and permanency

The obvious and pressing question of how permanent a reputation is - and whether it can be changed - is left unmentioned in the discussion of reputation in these chapters. Alongside, and mostly preceding, the swathes of pages dedicated to the study of reputation in IR for several decades, lies a treasure-trove of research on the permanency of reputation for an actor in labour economics. One key figure in this discussion who has, perhaps, done the most to reconcile how individuals and collectives are co-dependent on each other to maintain their reputational types, and how the old-addage from Warren Buffet that "It takes 20 years to build a reputation and five minutes to ruin it" rings mostly true, is Jean Tirole. Tirole (1996) put forward a theory on the micro-foundations of collective reputations.

Tirole suggests that reputation is a history-dependent concept, where administrations or organizations suffer from the sins of their elders for a long period of time after a negative shock to their reputation. This position is in opposition to the idea put forward in this dissertation; that reputations are group-based, not just dependent on history, and moreover, audience attention is critical to the relevancy of that reputation to how they perceive the actor. If the audience is

not primed to care about that reputational type, or another reputational type is more dominant, does it matter for their engagement with, or their perception of, that actor?

The findings here would suggest as attention wanes, or an actor can utilise covert strategies, or the saliency of the reputational type decreases, the damaged reputation is less meaningful. If there were a hierarchy of reputational types, then one could claim that one type of reputation *always* matters and therefore reputational damage to that type is long-lasting and persistent. This would be the argument for realists in reference to reputations for resolve, where the international system is anarchic and the balance of power precarious, or for governments and corruption, or for firms and product quality. However, the idea that damage to a reputation can be long-lasting because the individuals who belong to the collective face incentives to continue to cheat, defect, or shirk adds an interesting dimension that this dissertation does not cover. Future research could build on Tirole's insights, and the insights made in this dissertation, to evaluate how a one-time shock in reputation for an actor can affect engagement with that actor thereafter for certain types of reputations.

This type of reputational stickiness and the relationship between individual and collective reputations is a ripe area for future research. A good reputation allows governments of states to enter into international agreements that confer benefits, or engage in advantageous cooperative endeavours in the long run (Heymann, 1973, 822). Reputation has a similiar function for international organizations (IOs). The ability for IO's to solve collective action problems and remain focal points for state coordination is by maintaining their legitimacy (Tallberg & Zurn 2019, 582, Putnam 1988, Sommerer and Agné 2018). Agency

slack, through shirkage or slippage, can undermine the reputation of the IO and reduce their credibility. In a world of IO forum shopping, greater information transparency for domestic audiences, and high transaction costs of compliance, damage to the reputation of the IO can affect their competitive advantage as a site of state engagement (Morse and Keohane 2014; Zelli 2018). Cheating, defection, violating rules or norms, and, failing to fulfil or backtracking on commitments, can result in reputational damage. In such instances, bad reputations can be costly and difficult to repair. Future research could focus on the relationship between reputations of the members, and reputations of the collectives within an IO as a determinant of IO withdrawal/exclusion. Tirole's model might suggest that IO's move to exclude or distance member states following shocks to individual reputations to protect the collective reputation, while state's may move to withdraw from the IO following a shock to the collective reputation to protect their individual reputation. An interesting corollary offered by my dissertation findings is that withdrawals or exclusions could work through a cascade effect; as an IO suffers a reputational shock, and an early-mover withdraws, does this trigger a cascade of member withdrawal?

# 5 Reputation as a Group-Based Dynamic Concept

Observers do not make judgments on reputation based on an actor's characteristics or disposition alone. How could the observer say that the actions of the agent are suggestive of a good or bad type? How would they characterize, say, a state that had a cooperative reputation, versus a non-cooperative reputation? Or how would an audience perceive an agent to have a good reputation on climate change? Or, a human-rights abiding reputation? What is the threshold

of good versus bad behavior that leads the observer to believe the reputation is good or bad? There are two contributions I make to address these questions. The first is that audiences are diverse and exist in a group-dynamic, they will look to each other when forming judgments over an actor and coming to a conclusion over their reputational type. The second is that the reputations are interdependent, and the other members of the group within which an actor exists in a given situation will determine how observers cast judgments over the reputational type of the actor.

R. L. Jervis (1968, p. 11) touched on this idea of group-based reputations when he stated that, "To put this differently, although we may use the status quo as a reference point, it will be a status quo not in absolute terms but in relative or positional terms, as we look at whether we are maintaining our position vis-a-vis others." From this vantage point, reputation is a positional good where the rewards or benefits of maintaining a particular type of reputation are accrued not from how well the actor has performed in absolute terms, but rather the comparison of how the actor's behavior compares to their peers. Defining reputation as a dynamic concept that fluctuates in reaction to peer-behavior has two implications; reputation externalities for collective reputations, and reputation stealing externalities for individuals.

If you belong to a group, not by an act of consciously or officially joining necessarily, but by an association with the group or circumstantially you find yourself facing the same choice that others are also making, then the poor behavior of one member of the group can cast aspersions on other members of the group and lower overall group welfare. This is called a 'reputation externality'. An example of this in IR can be found when states which join together under

an organization or grouping find that the behavior of one agent can damage the reputation of the other agent. In order to preserve the reputation of the organization through which the organization derives its legitimacy, the organization will exclude those who threaten the collective's image. For example, the UNHRC prevented Russia from being re-elected to the organization in 2016, despite the fact that it had been a member since 2006, due to the human rights abuses in Ukraine and Syria. Similarly, following the tragic murder of journalist, Jamal Khashoggi, Saudi Arabia failed to secure their seat on the council. Or, perhaps, it can occur when heterogeneous agents or bodies develop common reputations through associations or long-standing alliances. On a more individual level, we know reputational externalities occur for leaders within political parties, where a leader or even a member of the party develops a bad reputation, they affect the reputation of the party as a whole. I found little evidence for reputational externalities in my dissertation when examining intra-industry network effects, indeed it seems more likely that this type of mechanism requires a tightly bound group of explicit members of the organization. Of course, the research design in this dissertation was not suited to this specific causal mechanism. Future research could use experimental survey designs of the public, or consumers, to detect whether the actions of one member of a group diminishes public perception of the collective. It may be the case that there is no real effect, but members themselves perceive and imagine this to be true.

The second effect is that when one member of the group deviates from the status quo, they reset the benchmark for all firms, leading to a decline in reputation (or loss of reputation) for agents that do not follow suit. When there is no crisis or no highly salient situation, we might say the reference point for judgments over an actor's behavior that are then used to determine the actor's

reputation is just the status quo. So, future behavior and/or decisions are understood in reference to fluctuating distributions of behavior and/or decisions. During an unprecedented event or crisis the situation may demand action that deviates from the norm. The new reference point or benchmark changes in reaction to the decisions and behavior of group members. Often, as is demonstrated in the findings from paper 2 and paper 3, this new reference point is determined by early-movers who ratchet up expectations on what is 'good' or 'bad' behavior. As in paper 2, all firms choosing to remain in Russia or pause operations, might lead audiences to expect that the firms themselves are unable to leave, or this is not feasible. In this instance, the reputation is not re-distributive, and there is no positional effect. When one or two firms leave, they send a strong signal that they are departing Russia because they no longer condone the behavior of Russia. In doing so they are saying that choosing to remain would be amoral. The result is that they shift the previous reference point of 'acceptable behavior' from staying to a new reference point; leaving. All firms that remain are now at risk of damage their reputational type. Similarly, when no firms disclose their behavior on climate change, or cooperate with voluntary regulatory bodies, the reference point for behavior is non-compliance and no reputations are at risk. When some firms begin to disclose, audiences are now aware that the new reference point for that reputation is to disclose; non-disclosure puts the firm at risk.

By discarding the traditional dispositional-situational dichotomy over the debate on whether reputation matters, and instead suggesting that judgments of an actor's disposition are only made when the observer is able to compare directly to the actor's peers or intra-group members, I also divorce reputation from necessarily having a hangover or spillover effect. By hangover I mean that the reputation built up in the moment will translate to future situations, for

instance Oye (1985, p. 16)'s idea that a reputation for cooperation in the moment will translate to a "lengthening of the shadow of the future". The degree of cooperativeness judged in a situation is dependent on the actor's peers. The spillover effect would suggest that an actor's reputational type built up in one situation might spill over to other similar situations. When reputation is considered group-based, this premise cannot hold, since other situations will require adjustments in which actors are involved (who are the peers or competitors), and what is the benchmark of behavior. The reputational type that an actor builds up in one arena (let's assume that the benchmark is for all actor's to comply since the majority of actor's are moving to a position of compliance), audience's will form judgments over the actor's reputation for failure to move to this position quickly. In another arena, the benchmark might be partial compliance, with few peer actor's moving to comply. Fudenberg and Tirole (1991, p. 391) indicate that this is where reputational types do not need to be fixed, "an open question in this field is what happens when the incumbent's type need not be the same in each contest, so that the incumbent can be tough in some contests and weak in others".

The idea of multiple audiences and relative reputations espoused in this dissertation harks back to an idea proposed by Fearon (1994, p. 577) who discusses *relative audience costs* - "relative audience costs matter: the side with a stronger domestic audience (e.g., a democracy) is always less likely to back down than the side less able to generate audience costs (a nondemocracy)". In this context Fearon is alluding to the idea that autocratic leaders are insulated from audience costs, and given this, reputational costs vary between regime types. What this idea also suggests, however, is that reputational types can be relative depending on the strength of the audience. Which is to say when an actor faces

two audiences simultaneously, with different expectations over reputation or a conflicting reputational type, the relative strength of one audience over another may determine how that actor makes a decision over which reputational type to signal. Paper two suggest that actor's employ specific strategies to manage both audiences. However, another avenue would be for the actor to satisfy the strongest audience. Which audience is strongest or most critical may not always be consistent.

# 6 On dueling Audiences and Conflicting Reputations

A key finding in this dissertation that challenges the classical understanding of the relationship between the bearer of a reputation and the observer is that there can exist multiple and conflicting reputations among multiple and conflicting groups of observers. The crisis bargaining model of reputation, and reputational audience cost models used in the literature on resolve (Fearon 1994; Schelling 1966), in some cases do bring into question multiple audiences; for instance the domestic versus the international (Putnam 2017). What this literature does, however, is hold constant the reputational type, where the actor may gain from signaling a reputational type among one audience, while they same signal will cost them (Kertzer and Brutger 2016; Casler, Clark, and Zucker 2024; Brutger and Kertzer 2018). The invoking of multiple audiences is also usually across points of time; the costs of backing down for your audience now will cost you with that same audience in the future (i.e. resolve as a form of deterrence) (Huth, Bennett, and Gelpi 1992; Grieco 2001; Walter 2006). Clare and Danilovic (2010) also invokes the concept of multiple audiences, but again, the reputational type is held constant, and the audiences themselves are essentially the same types of

actors - the idea is that rather than an actor being incentivized to show resolve for future acts of aggression, they're incentivized when facing multiple rivals for contemporary deterrence against those rivals.

In the framework of reputation as a group-based concept put forward here, actors face multiple audiences with conflicting expectations over that actor's reputational type. Actors' therefore, are not just facing audience costs from failing to signal one type of reputation among one audience; any decision to signal a reputational type could automatically incur reputational gains and losses. This consideration therefore changes how actor's behave and respond to concern for reputation and audience costs. Previous work seeking to measure reputational costs has suggested that actors are insensitive to reputational costs if they fail to signal that type (Weeks 2008); however, by introducing the assumption of multiple audiences, an actor's behavior may appear more like reputational management. The knowledge of multiple audiences can also suggest that some actor's may be more likely to employ covert and overt strategies to satisfy both audiences at once; this is particularly true of special interests. An actor may also rely on low informational transparency in order to hide or obfuscate among another audience.

This dissertation discusses the role of dueling reputations and conflicting audiences, however it does not explore how individuals might perceive different signals of different reputations under different situations. Future research could examine which signals observers are more inclined to interpret for different types of reputations. Situations conferring inferences over reputation in world politics - conflicts, international agreements and alliances - take place in diverse informational environments which may place different weights on particular in-

dicators. Certain types of actors and leaders are more visible in certain types of settings (i.e. conflict situations and leaders), where signaling reputations for resolve is necessary. However, observers perceive reputations for cooperation in situations involving economic alliances or bilateral loans. Here indicators of risk are most critical: i.e. past behavior (committing to agreements or timely repayments). Thus, observers use different indicators to assess different types of reputations. Observers will use the most relevant indicator as a heuristic short-cut when assessing a reputational type in a given situation; meaning the weight of an indicator is conditional on the reputational type. This would be a fruitful avenue for future research.

Another avenue for research emerging from the idea of dueling reputations is whether observers value some reputational types over others. Scholars in IR have found evidence that states and leaders can acquire reputations for various qualities including resolve, honesty, violence or hostility, keeping alliance commitments, repaying debts, or even handling secessionist groups (Guisinger and Smith 2002; Gibler 2008; Tomz 2007; Walter 2009; Peterson 2013). Just as is the case in consumer economics, some types of reputations are more fungible than others. Where quality matters most for the consumer, the literature in IR has focused on three distinctive reputations that encompass more narrower variations; 'resolve', 'cooperation ', and 'compliance'.[3] Each of these reputations confers important information to the reputation-observer about how they should engage with the reputation-holder in a variety of situations or contexts. What's unclear here, is why some states would prioritise particular reputations - or indeed, why observers would value one type of reputation over the other. I

[3]Credibility might be considered a reputational type, however the literature has thus far treated credibility as distinctive from reputation (signifying reliability, or perceived likelihood an agent will meet commitments) where reputation is an ingredient of credibility. Credibility also overlaps with resolve, but the latter is a narrower concept.

have suggested that this is the result of certain crises that prime audiences to care about particular reputations; however in the absence of crisis, variation in which type of reputation matters would be an interesting line of research. While this dissertation has shown conflicting reputations exist, the micro-foundations of why this is the case is left for future research.

Another avenue is vantage points of audience members. Individuals make judgments over reputation because the actions of an agent in a given situation has inter-temporal consequences; if an agent defects on an alliance observers may infer this behavior is typical of the agent which has consequence for future alliances. But risk is not assessed evenly and observers differ in their risk vantage point. It may be the case that individuals located in higher-threat contexts are more likely to privilege reputations for resolve, while individuals in lower-threat contexts are likely to privilege cooperative reputations. This could also extend to firms; perhaps consumers value reputations for environmentalism during periods of high climate volatility, or when citizens are primed to consider climate, but during periods of poor economic growth, or high inflation, consumers may value reputations for brand quality, or price (i.e. the trade-off between the firm signaling their lower prices versus economic costs). Preliminary evidence already suggests that consumers vary in who is willing to pay more for environmentally-friendly goods.

Focusing on audience members, it would be interesting to extend the findings on when consumers care about environmentally-conscious reputations of firms, and which/when firms feel pressured to send this signal through disclosing their climate-conscious actions. It would be interesting to compare across country contexts, where audiences may value different types of reputation depending

on their socio-economic status, their level of climate-consciousness, internet-penetration levels and therefore transparency of behavior, and the cultural norms around Corporate Social Responsibility (CSR). Since awareness of reputation among other agents affects how they engage with an agent, reputation can become an enforcement mechanism to shape pro-social behavior. But social preferences - and what is considered pro-social behavior - are often endogenous to particular societal environments (Bowles 1998). A burgeoning literature uncovered the connection between particular demographic features and social preferences. For instance, Eckel and Grossman (2003) and Andreoni and Vesterlund (2001) found a link between gender and altruism in dictator game experiments, with Croson and Gneezy (2009) finding that women are more primed by concern for reputation. In public good games, List (2004) found that older subjects contributed more of their endowment to the public good than younger participants. Geographical feature may also influence divergent pro-social behavior. Countries with large urban-rural gaps could dilute the effects of a 'reputational mechanism', since social norm enforcement is lower with less monitoring (Carpenter 2007), and rural subjects internalize these norms of cooperation.

Beyond demographic or geographic features, societal features specific to a country can also shape pro-social behavior. Public good games conducted in advanced democracies like the United States or Switzerland find that individuals will often punish free-riders who fail to contribute to the public good, inducing cooperative pro-social behavior in future games (Fehr and Gächter 2000; Fehr and Gächter 2002; Gächter and Herrmann 2009). However, when these same experiments were carried out in Russia, Gächter and Herrmann (2011) found that punishment did not have a cooperation-enhancing effect and subjects would punish other subjects who contributed the same or more than the punishing

subject (misdirected punishment). One line of reasoning has explained this by looking at variation in levels of trust among societies; punishment promotes cooperation in high-trust societies where individuals count on each other to punish free-riders and promote the norm cooperation, and feelings of guilt for not contributing are more pronounced (Balliet and Van Lange 2013b; Balliet and Van Lange 2013a). Where these norms are less shared or enforced (low trust societies), punishment induces anger and retaliation instead (Gintis 2008). This level of variation in pro-social behavior should also affect how individuals perceive collective reputations; the type of reputation an actor has in a particular domestic or international market, as well as the attributes that matter most in forming a desirable reputation. I leave this line of inquiry to future research.

Recent work has suggested the reputational mechanism driving audience costs have decomposed into a double-barrelled cost; 1) leaders that threaten to use force, and 2) leaders that are inconsistent in following through with their threat to use force. Although the conclusion derived from this was evidence that "not all audiences are concerned about reputational costs" (Kertzer and Brutger 2016, p. 244), with further research suggesting that there may be a gendered split in who punishes for force versus consistency (Barnhart et al. 2020), what this really suggests is that individuals care about different types of reputations. It would be interesting to extend the so-called double-barreled cost to other types of reputation in other situations; for instance, do people reward firms for their initial posturing on climate change, and punish for lack of follow-through? Are companies that signal their intention to leave Russia, but then choose to remain, punished by consumers or stakeholders? Some of these questions require more granular data, or experimental survey data, to uncover.

Another line of inquiry pertains to whether situations themselves form focal points that guide behavior for reputations to emerge. As has been repeatedly stated here, reputations are rarely formed in vacuums. Despite the lively debate over whether dispositional traits (where the nature or character of an individual determines an action) or situational traits (whereby most agents in the same situation behave similarly) matter more (Mercer 2018), what is more likely is that dispositional traits interact with situational characteristics in the eye of the beholder. Situations form 'focal points' or scope conditions that shape incentives to convey particular reputations, or prompt the observation of particular reputations (Kreps 1990). For cooperative situations, where the context is characterized by repeated interaction over time, trust between agents, reciprocity, compromise, and negotiation, an agent's dispositional traits related to these situational characteristics will be given greater weight by observers. In contrast, in conflict disputes, crisis bargaining or security arrangements, agents are primed to project traits for resolve, and observers to seek out these cues (Kahneman, Sibony, and Sunstein 2021). These situations are characterized by signaling military strength, consistency of follow-through (Fearon 1994; Tomz 2007; Weeks 2008; Levendusky and Horowitz 2012; Levy et al. 2015; Lin-Greenberg 2019), and state-level capacity to respond.

Exploring how audience observers assess reputational types prompts a related question: do observers evaluate reputation differently when it's their own state versus a foreign state – or whether it's a domestic firm versus a foreign firm? Studies on resolve suggest that states that back down after issuing threats face external costs on reputations for resolve as perceived by third-party countries. But these reputational costs are also internalised (Luo 2022). For instance, J. C. Weiss and Wallace (2021) argue that the Chinese Communist Party's

regime survival is linked to Beijing's aggressive stance on maritime and island disputes. Similarly, Akio (2017) argues that the stability of the CCP's power-base will determine whether the state takes a hard-line approach to Japan in the future. But these observers have in mind their own country as the subject of reputation; do observers in these countries have different reference points for reputational types?

Other work would suggest a level of heterogeneity in how observers evaluate their home state versus a foreign state's reputation. While citizens often reward leaders for projecting a reputation for resolve through aggression and consistency (J. Weiss and Dafoe 2018; Oakes 2006; Fearon 1994; Tomz 2007; Snyder 1991; Clary, Lalwani, and Siddiqui 2021), when citizens are asked to evaluate foreign policy towards other states, particular cooperative characteristics are viewed favourably (Bell and Quek 2015). Drawing on insights related to conditional cooperation and reciprocity (Tingley and Tomz 2014; Kertzer and Rathbun 2015; Rabin 2002), one study suggested that moral sentiments around fairness and reciprocity were strong predictors of attitudes towards international relations. Kertzer and Rathbun (2015) and Chilton (2016) found evidence suggesting public support for compliance with international law during interstate disputes was driven in part by views on reciprocity. The combination of research on outward-attitudes towards foreign states and cooperation with insights on resolve-orientated public inward-attitudes toward home-state governments, suggests individuals will be more favourable of *foreign state* reputations for cooperation.[4]

In summation, this dissertation is attempting to pave new in-roads in which

[4] Although not the focus of this paper, the nature of the relationship between observers and types of states (allies or adversaries) undoubtedly conditions perceptions of reputation (Press 2005).

to understand reputation in the context of many areas of IR. It invites future researchers to continue some of the tangents and lines of inquiry created by the puzzles this dissertation engages with.

# References


Akio, Takahara (2017). "Forty-four years of Sino–Japanese diplomatic relations since normalization". In: *China-Japan relations in the 21st century: Antagonism despite interdependency*, pp. 25–65.

Andreoni, James and Lise Vesterlund (2001). "Which is the fair sex? Gender differences in altruism". In: *The Quarterly Journal of Economics* 116.1, pp. 293–312.

Balliet, Daniel and Paul AM Van Lange (2013a). "Trust, conflict, and cooperation: a meta-analysis." In: *Psychological bulletin* 139.5, p. 1090.

— (2013b). "Trust, punishment, and cooperation across 18 societies: A meta-analysis". In: *Perspectives on Psychological Science* 8.4, pp. 363–379.

Barnhart, Joslyn N et al. (2020). "The suffragist peace". In: *International Organization* 74.4, pp. 633–670.

Bell, Mark and Kai Quek (2015). "Realism, Idealism, and American Public Opinion on Nuclear Disarmament". In.

Bowles, Samuel (1998). "Endogenous preferences: The cultural consequences of markets and other economic institutions". In: *Journal of economic literature* 36.1, pp. 75–111.

Brutger, Ryan and Joshua D Kertzer (2018). "A dispositional theory of reputation costs". In: *International Organization* 72.3, pp. 693–724.

Carpenter, R Charli (2007). "Setting the advocacy agenda: Theorizing issue emergence and nonemergence in transnational advocacy networks". In: *International Studies Quarterly* 51.1, pp. 99–120.

Casler, Don, Richard Clark, and Noah Zucker (2024). "Do Pledges Bind? The Mass Politics of International Climate Targets". In.

Chilton, Adam S (2016). "The political motivations of the United States' bilateral investment treaty program". In: *Review of International Political Economy* 23.4, pp. 614–642.

Clare, Joe and Vesna Danilovic (2010). "Multiple audiences and reputation building in international conflicts". In: *Journal of Conflict Resolution* 54.6, pp. 860–882.

Clary, Christopher, Sameer Lalwani, and Niloufer Siddiqui (2021). "Public opinion and crisis behavior in a nuclearized south Asia". In: *International Studies Quarterly* 65.4, pp. 1064–1076.

Coleman, James S (1994). *Foundations of social theory*. Harvard university press.

Croson, Rachel and Uri Gneezy (2009). "Gender differences in preferences". In: *Journal of Economic literature* 47.2, pp. 448–474.

Dafoe, Allan, Jonathan Renshon, and Paul Huth (2014). "Reputation and status as motives for war". In: *Annual Review of Political Science* 17.1, pp. 371–393.

De Mesquita, Bruce Bueno et al. (2004). "Testing novel implications from the selectorate theory of war". In: *World Politics* 56.3, pp. 363–388.

Eckel, Catherine C and Philip J Grossman (2003). "Forecasting risk attitudes: An experimental study of actual and forecast risk attitudes of women and men". In: *Virginia Tech Department of Economics Working Paper*.

Fearon, James D (1994). "Domestic political audiences and the escalation of international disputes". In: *American political science review* 88.3, pp. 577–592.

Fehr, Ernst and Simon Gächter (2000). "Cooperation and punishment in public goods experiments". In: *American Economic Review* 90.4, pp. 980–994.

Fehr, Ernst and Simon Gächter (2002). "Altruistic punishment in humans". In: *Nature* 415.6868, pp. 137–140.

Fombrun, Charles and Mark Shanley (1990). "What's in a name? Reputation building and corporate strategy". In: *Academy of management Journal* 33.2, pp. 233–258.

Fudenberg, Drew and Jean Tirole (1991). *Game theory*. MIT press.

Gächter, Simon and Benedikt Herrmann (2009). "Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1518, pp. 791–806.

— (2011). "The limits of self-governance when cooperators get punished: Experimental evidence from urban and rural Russia". In: *European economic review* 55.2, pp. 193–210.

Gibler, Douglas M (2008). "The costs of reneging: Reputation and alliance formation". In: *Journal of Conflict Resolution* 52.3, pp. 426–454.

Gintis, Herbert (2008). "Punishment and cooperation". In: *Science* 319.5868, pp. 1345–1346.

Grieco, Joseph M (2001). "Repetitive military challenges and recurrent international conflicts, 1918–1994". In: *International Studies Quarterly* 45.2, pp. 295–316.

Guisinger, Alexandra and Alastair Smith (2002). "Honest threats: The interaction of reputation and political institutions in international crises". In: *Journal of Conflict Resolution* 46.2, pp. 175–200.

Hobbes, Thomas (n.d.). "Leviathan (Hobbes book)". In: ().

Huth, Paul, D Scott Bennett, and Christopher Gelpi (1992). "System uncertainty, risk propensity, and international conflict among the great powers". In: *Journal of Conflict Resolution* 36.3, pp. 478–517.

Jervis, Robert (1989). *The logic of images in international relations*. Columbia University Press.

— (2017). *How statesmen think: The psychology of international politics*. Princeton University Press.

Jervis, Robert Louis (1968). *Signals, Indices, and the Projection and Manipulation of Images in International Relations*. University of California, Berkeley.

Kahneman, Daniel, Jack L Knetsch, Richard Thaler, et al. (1986). "Fairness as a constraint on profit seeking: Entitlements in the market". In: *American economic review* 76.4, pp. 728–741.

Kahneman, Daniel, Olivier Sibony, and Cass R Sunstein (2021). *Noise: a flaw in human judgment*. Hachette UK.

Kertzer, Joshua D and Ryan Brutger (2016). "Decomposing audience costs: Bringing the audience back into audience cost theory". In: *American Journal of Political Science* 60.1, pp. 234–249.

Kertzer, Joshua D and Brian C Rathbun (2015). "Fair is fair: Social preferences and reciprocity in international politics". In: *World Politics* 67.4, pp. 613–655.

Kreps, David M (1990). *Game theory and economic modelling*. Oxford University Press.

Kuran, Timur and Cass R Sunstein (1998). "Availability cascades and risk regulation". In: *Stan. L. Rev.* 51, p. 683.

Lake, David A (2013). "Theory is dead, long live theory: The end of the Great Debates and the rise of eclecticism in International Relations". In: *European journal of international relations* 19.3, pp. 567–587.

Levendusky, Matthew S and Michael C Horowitz (2012). "When backing down is the right decision: Partisanship, new information, and audience costs". In: *The Journal of Politics* 74.2, pp. 323–338.

Levy, Jack S et al. (2015). "Backing out or backing in? Commitment and consistency in audience costs theory". In: *American Journal of Political Science* 59.4, pp. 988–1001.

Lin-Greenberg, Erik (2019). "Backing up, not backing down: Mitigating audience costs through policy substitution". In: *Journal of Peace Research* 56.4, pp. 559–574.

List, John A (2004). "Young, selfish and male: Field evidence of social preferences". In: *The Economic Journal* 114.492, pp. 121–149.

Luo, Shuxian (2022). "The rising power's audiences and cost trade-offs: explaining China's escalation and deescalation in maritime disputes". In: *Asian Security* 18.2, pp. 172–199.

Mercer, Jonathan (2010). *Reputation and international politics*. Cornell University Press.

Milinski, Manfred, Dirk Semmann, and Hans-Jürgen Krambeck (2002). "Reputation helps solve the 'tragedy of the commons'". In: *Nature* 415.6870, pp. 424–426.

Oakes, Amy (2006). "Diversionary war and Argentina's invasion of the Falkland Islands". In: *Security Studies* 15.3, pp. 431–463.

Oye, Kenneth A (1985). "Explaining cooperation under anarchy: Hypotheses and strategies". In: *World politics* 38.1, pp. 1–24.

Peterson, Timothy M (2013). "Sending a message: The reputation effect of US sanction threat behavior". In: *International Studies Quarterly* 57.4, pp. 672–682.

Plutarch, Scott-Kilvert et al. (1960). "The rise and fall of Athens: nine Greek lives". In: *(Penguin)*.

Potter, Philip BK and Matthew A Baum (2014). "Looking for audience costs in all the wrong places: Electoral institutions, media access, and democratic constraint". In: *The Journal of Politics* 76.1, pp. 167–181.

Press, Daryl Grayson (2005). *Calculating credibility: How leaders assess military threats*. Cornell University Press.

Putnam, Robert D (2017). "Diplomacy and domestic politics: the logic of two-level games". In: *International organization*. Routledge, pp. 437–470.

Rabin, Matthew (2002). "Inference by believers in the law of small numbers". In: *The Quarterly Journal of Economics* 117.3, pp. 775–816.

Renshon, Jonathan, Keren Yarhi-Milo, and Joshua D Kertzer (2023). "Democratic reputations in crises and war". In: *The Journal of Politics* 85.1, pp. 1–18.

Roberts, Peter W and Grahame R Dowling (2002). "Corporate reputation and sustained superior financial performance". In: *Strategic management journal* 23.12, pp. 1077–1093.

Schelling, Thomas C (1966). "Arms and influence Yale University press". In: *New Haven*.

— (2001). "The strategy of conflict". In: *Readings in Games and Information* 3, p. 83.

Slantchev, Branislav L (2012). "Audience cost theory and its audiences". In: *Security Studies* 21.3, pp. 376–382.

Snyder, Glenn H (1991). "Alliances, balance, and stability". In: *International Organization* 45.1, pp. 121–142.

Sunstein, Cass R (2019). *How change happens*. Mit Press.

Tingley, Dustin and Michael Tomz (2014). "Conditional cooperation and climate change". In: *Comparative Political Studies* 47.3, pp. 344–368.

Tirole, Jean (1996). "A theory of collective reputations (with applications to the persistence of corruption and to firm quality)". In: *The Review of Economic Studies* 63.1, pp. 1–22.

Tomz, Michael (2007). "Domestic audience costs in international relations: An experimental approach". In: *International Organization* 61.4, pp. 821–840.

Walter, Barbara F (2006). "Building reputation: Why governments fight some separatists but not others". In: *American Journal of Political Science* 50.2, pp. 313–330.

— (2009). *Reputation and civil war: Why separatist conflicts are so violent*. Cambridge University Press.

Weeks, Jessica L (2008). "Autocratic audience costs: Regime type and signaling resolve". In: *International Organization* 62.1, pp. 35–64.

— (2012). "Strongmen and straw men: Authoritarian regimes and the initiation of international conflict". In: *American Political Science Review* 106.2, pp. 326–347.

Weisiger, Alex and Keren Yarhi-Milo (2015). "Revisiting reputation: How past actions matter in international politics". In: *International Organization* 69.2, pp. 473–495.

Weiss, Jessica and Allan Dafoe (2018). "Authoritarian audiences and government rhetoric in international crises: Evidence from China". In: *International Studies Quarterly*.

Weiss, Jessica Chen and Jeremy L Wallace (2021). "Domestic politics, China's rise, and the future of the liberal international order". In: *International Organization* 75.2, pp. 635–664.

Yarhi-Milo, Keren (2018). *Who fights for reputation: The psychology of leaders in international conflict*. Vol. 156. Princeton University Press.